Education and Training in Autism and Developmental Disabilities

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

Volume 51 Number 1
March 2016
Education and Training in Autism and Developmental Disabilities
The Journal of the Division on Autism and Developmental Disabilities,
The Council for Exceptional Children

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

Consulting Editors
Martin Agran
Kevin Ayres
Devender Banda
Laura Bassette
Emily Bouch
Amanda Bourot
Michael P. Brady
Stacy Carter
David Cihak
Ginevra Courtade
Samuel A. DiGangi
Teresa Doughby
Jennifer Ganz
Juliet Hart Barnett
Carolyn Hughes
Bree Jimenez
Russ Lang
Justin Leaf
Wendy Machalicek
Michael Mayton

Editorial Assistant: Jessica Holloway-Libell
Arizona State University
Mary Lou Fulton Teachers College

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. Membership is on a yearly basis. All inquiries concerning membership, subscription, advertising, etc. should be sent to the Division on Autism and Developmental Disabilities, 2900 Crystal Drive, Suite 1000, Arlington, VA 22202-3557. Advertising rates are available upon request.

Manuscripts should be typed, double spaced, and sent (three copies) 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 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, PsykolNFO, 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 Dianne Zager
PRESIDENT David Cihak
PRESIDENT-ELECT Elizabeth West
VICE PRESIDENT Jordan Shurr
SECRETARY Meaghan McCollow
TREASURER Gardner Umbarger

Members

Debra Cote
Richard Gariguio
Beth Kavanaugh
Jenny Root (Student Representative)
Leah Wood

Executive Director
Teresa Doughty
Publications Chair
Michael Wehmeyer
Communications Chair
Emily C. Bouch
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, 2900 Crystal Drive, Suite 1000, Arlington, Virginia 22202-3557. 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, 2900 Crystal Drive, Suite 1000, Arlington, Virginia 22202-3557.
Economic and Demographic Factors Impacting Placement of Students with Autism
JENNIFER A. KURTH, ANN M. MASTERGEORGE, and KATHERINE PASCHALL

Meta-Analysis of Pivotal Response Training for Children with Autism Spectrum Disorder
GULDEN BOZKUS-GENC and SERIFE YUCESOY-OZKAN

An Evaluation of App-Based and Paper-Based Number Lines for Teaching Number Comparison
PEI-LIN WENG and EMILY C. BOUCK

Embedded Simultaneous Prompting Procedure to Teach STEM Content to High School Students with Moderate Disabilities in an Inclusive Setting
SARA HEINRICH, BELVA C. COLLINS, VICTORIA KNIGHT, and AMY D. SPRIGGS

An Evaluation of Constant Time Delay and Simultaneous Prompting Procedures in Skill Acquisition for Young Children with Autism
JULIE A. ACKERLUND BRANDT, SARA WEINKAUF, NICOLE ZEUG, and KEVIN P. KLATT

Family Generated and Delivered Social Story Intervention: Acquisition, Maintenance, and Generalization of Social Skills in Youths with ASD
SERAY OLÇAY-GÜL and ELIF TEKIN-İFTAR

Video Modeling to Teach Social Safety Skills to Young Adults with Intellectual Disability
CORRINE E. SPIVEY and LINDA C. MECHLING

Differentiated Effects of Sensory Activities as Abolishing Operations via Non-contingent Reinforcement on Academic and Aberrant Behavior
G. RICHMOND MANCIL, TODD HAYDON, and MARTY BOMAN

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

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

June 2016

Effects of a peer-mediated intervention on social interactions of students with low-functioning autism and perceptions of typical peers. Lisa A. Simpson and Yvonne Bui, Connie L. Lurie College of Education, One Washington Square, San Jose State University, San Jose, CA 95192-0078.

Sequence learning with stochastic feedback in a cross-cultural sample of boys in the autistic spectrum. Maren Hentschel, Christiane Lange-Küttner, and Bruno B. Averbeck, London Metropolitan University, Tower Building T6-20, 166-220 Holloway Road, Tower Building T6-20, London, N7 8DB, UK.

Modifications of the one-more-than technique: A comparison of two strategies for teaching purchasing skills to students with intellectual disability in Taiwan. Guo-Liang Hsu, Jung-Chang Tang, Wu-Yuin Hwang, Yung-Chang Li, Jung-Chao Hung, and Chun-Hwa Wei, National Central University, No. 300, Jhongda Rd., Jhongli City, Taoyuan County, 3201, TAIWAN.

Effects of computer-based video instruction on the acquisition and generalization of grocery purchasing skills for students with intellectual disabilities. Minkowan Goo, William J. Therrien, and Youjia Hua, Texas Women’s University, Teacher Education, P.O. Box 425769, Denton, TX 76204-5769.

Use of a behavioral art program to improve social skills of two children with autism spectrum disorders. Wan-Chi Chou, Gabrielle Lee, and Hua Feng, 5528 Whitfield Drive, Troy, MI 48098.

Parent implementation of RECALL: A systematic case study. Kelly Whalon, Mary Frances Hanline, and Jackie Davis, Florida State University, 1114 West Call Street, Tallahassee, FL 32307.

Effects of peer-mediated literacy based behavioral intervention on the acquisition and maintenance of daily living skills in adolescents with moderate to severe disabilities. Michael Brady, Christine Honsberger, Jessica Cadette, and Toby Honsberger, Exceptional Student Education, Florida Atlantic University, 777 Glades Road, Boca Raton, FL 33431.

Effectiveness and acceptability of parent-implemented behavior interventions for children with autism in three African American families. Rachel E. Robertson, Department of Instruction and Learning, University of Pittsburgh, 5146 WWPH, 230 South Bouquet Street, Pittsburgh, PA 15260.

Address is supplied for author in boldface type.
Economic and Demographic Factors Impacting Placement of Students with Autism

Jennifer A. Kurth
University of Kansas

Ann M. Mastergeorge and Katherine Paschall
University of Arizona

Abstract: Educational placement of students with autism is often associated with child factors, such as IQ and communication skills. However, variability in placement patterns across states suggests that other factors are at play. This study used hierarchical cluster analysis techniques to identify demographic, economic, and educational covariates associated with placement patterns across states in highly inclusive, moderately inclusive, moderately restrictive, and highly restrictive clusters. Findings indicate that highly inclusive states are more rural, have more adults with high school diplomas and more White citizens compared to other clusters. States that are highly restrictive were largely less economically and racially privileged. These findings suggest an inequitable access to the least restrictive environment for students with autism. Implications of these findings are included.

There is an increasing number of students with autism spectrum disorders (ASD) being identified (Baio, 2012) and receiving special education services in U.S. schools (U.S. Department of Education, 2008). As students receive an autism diagnosis and enter schools, educational teams must determine the appropriate manner and placement for their education (Individuals with Disabilities Education Improvement Act [IDEA], 2004). Educational teams consist of invested individuals, including special and general education teachers, parents, administrators, school psychologists, and other education professionals (e.g., speech-language pathologists) who determine eligibility for special education, individual goals and services, and the settings in which those goals and services will be delivered (IDEA, 2004). Placement decisions involve deciding in which setting individual goals and services will be delivered and the amount of time (typically expressed as a percentage of time) in which students with ASD will be educated in the general education setting.

Educational teams tasked with making placement decisions for students with ASD arrive at their decisions for a variety of reasons, including an analysis of factors that are specific to a child (e.g., cognitive ability and social skills) and factors that are external to the child (e.g., locally available resources). While child factors (e.g., age, IQ, and skills) are often assumed to be primary determinants of placement decisions, and likely reflect the intent of IDEA to focus on unique child needs, state of residence has emerged as an important factor in educational placement. In fact, variability of placement by state or geographic region has been associated with placement patterns for a number of disability categories, including autism (Kurth, 2014), learning disability (McLeskey, Landers, Hoppey, & Williamson, 2011), intellectual disability (Katsiyannis, Zhang, & Archwmeny, 2002), and emotional behavioral disorders (Landrum, Katsiyannis, & Archwmeny, 2004). The fact that state of residence is an enduring factor in determining placement decisions is a strong indicator that child-specific factors alone do not account for placement decisions. Instead, this variability suggests that there are important factors within and across U.S. states impacting placement decisions. Because placement decisions have enduring ramifications on student academic out-
comes (e.g., Kurth & Masters, 2010, 2012) and, because once placed in a particular educational environment, students rarely leave that type of setting (White, Scahill, Klin, Koenig, & Volkmar, 2007), these placement decisions have critically important lifelong impacts on students (Test et al., 2009).

While IDEA requires that schools place students in the least restrictive environment to meet their needs, this mandate has often been misinterpreted as a need to provide a continuum of placement options (Taylor, 1988). A placement is considered less restrictive when students in that placement have more access to the general curriculum and setting; it is considered more restrictive when students have limited access to the general curriculum and setting. Placement of students with ASD in less restrictive settings has been associated with academic learning (e.g., Kurth & Mastergeorge, 2010) and social engagement (e.g., Lyons, Cappadocia, & Weiss, 2011), although access to less restrictive settings is unequal for students from varying backgrounds.

Specifically, previous research has examined student-level, family-level, and social-level factors associated with restrictiveness of special education placement. Students from high-poverty schools, as well as those representing racially minoritized groups, are more likely to be placed in more restrictive special education placements compared to White students and students from higher socioeconomic backgrounds (Fierros & Conroy, 2002). Further, Cosier and Causton-Theoharis (2011) used hierarchical regression techniques to assess the extent to which various economic and demographic variables predict the level of student participation in inclusive settings (defined as 80% or more of the school day in general education settings) in the state of New York. These authors found inclusive education was positively associated with higher per pupil spending on general education students, less per pupil spending on special education students, and was negatively associated with percentage of students receiving a free and reduced lunch.

Given the variability in placement patterns for students with ASD, and the lack of guidance from IDEA related to how students with disabilities are referred, evaluated, and placed in special education (Donovan & Cross, 2002), the aims of this study are to (1) profile state placement patterns for students with ASD into more or less restrictive placements and (2) examine within- and across-state covariates that may explain patterns of restrictiveness of placement of students with ASD.

**Method**

The federal government requires states to monitor the implementation of IDEA with the intent of improving educational results and functional outcomes for students with disabilities. One component of the monitoring approach consists of 20 IDEA Part B indicators (OSEP, 2009), including Indicator 5, which measures participation of students with disabilities in general education settings (least restrictive environment, or LRE). Indicator 5 requires states to report the percentage of students ages 6–21 served in the following three categories:

- **Category A:** Inside the regular class 80% or more of the day
- **Category B:** Inside the regular class less than 40% of the day, and
- **Category C:** Educated in separate schools, residential facilities, or homebound/hospital placements

**Inclusive Education**

Inclusive education is defined as a community of belonging (Artiles & Kozleski, 2007) where students have supports provided to address their needs. Inclusive education may be further defined as the provision of the range of supports and services provided to students in general settings to meet their unique learning needs thus maximizing student learning and participation. State level data prevents analysis of the types of supports and community developed within classrooms, and therefore percentage of time is used a proxy measure of inclusivity (Cosier & Causton-Theoharis, 2011). For our purposes here, a placement is considered “highly inclusive” when students spend 80% or more of their school day in general education settings (Category A). A placement is considered “moderately inclusive/restrictive” when students with ASD visit a general education classroom for...
portions of the school day but receive the majority of their education in a separate setting (Category B). Finally, a placement is considered “highly restrictive” when students with ASD are educated in separate schools or facilities (Category C).

Covariates

The covariates in this analysis include both economic, demographic, and disability status data from the 50 U.S. states and District of Columbia (referred to hereafter as a ‘state’). Economic variables include: (a) percent of population within a state living below poverty (as defined by the U.S. Census Bureau); (b) percent of students receiving a free or reduced lunch in the state; (c) median household income within a state; and (d) per pupil spending in the state. For this analysis, we were not able to obtain reliable data (e.g., per pupil spending) for each state related to only students with ASD, thus these variables reflect all students in the state. Demographic variables include: (a) percent of population living in an urban area (as defined by U.S. Census Bureau); (b) percent of people in the state aged 25 and older who have a high school diploma; (c) percent of people in the state aged 25 and older who have a bachelor’s degree; (d) race/ethnicity; and (e) language spoken at home. Disability status variables include: (a) percent of all IDEA eligible students with ASD in a state; (b) the number of general education students for every one student with ASD in the state; (c) number of students with IEPs in a state; and (d) number of students with ASD in the state.

Data and Sampling

Students with ASD, ages 6–22, in the 51 U.S. states were included in this analysis for the year 2012. Defining students with ASD is complicated by differences between clinical definitions of ASD (from the Diagnostic and Statistical Manual, DSM-V; American Psychological Association, 2013) and administrative definitions of ASD (from IDEA); further complicating matters, there is across state variability in administrative definitions of ASD (Travers, Krzemien, Mulcahy, & Tincani, 2014). Considering these challenges, it is likely that discrepancies exist; however, these data are deemed the best available at this time (e.g., Kurth, Morningstar, & Kozleski, 2014).

Three publicly available data sources were used for this analysis. Data on placement by disability label was obtained from the Office of Special Education Programs (OSEP) at www.idea-data.org. These data are collected and reported annually to OSEP by each state. Total child count for all disability categories for students ages 6 to 22, and total child count for students with ASD ages 6–22, in all U.S. states and Washington DC for the year 2012 were analyzed to determine the proportion of students with ASD within each state. Demographic information for each state (race/ethnicity, language spoken at home, high school diploma rate, bachelor’s degree rate, median household income, percent of persons living below poverty, and percent of the population living in urban areas) was obtained from the U.S. Department of the Census at http://quickfacts.census.gov/qfd/states. Finally, information on per pupil spending (for all students), and total numbers of students grades K-12 in each state was obtained from the National Center for Education Statistics (www.nces.ed.gov). Data tables were downloaded from these sources and copied into an SPSS 21.0 worksheet for analysis.

Data Analysis

Statistical analyses were conducted in SPSS 21.0. We employed hierarchical cluster analysis using Ward’s method and the squared Euclidian distance to determine clusters of students in each placement category (A-C). One-way ANOVAs, along with a post-hoc Tukey’s test, were utilized to validate the presence of unique clusters within the dataset. In addition, ANOVAs were computed to determine if the clusters differed significantly from each other on state demographic, economic, and disability status characteristics.

Person-centered analyses allow for the estimation of distinct, homogeneous subgroups. These subgroups can then be compared on a variety of covariates. Empirically derived person-centered analyses supplement the research in this field, because they move beyond the Census Bureau’s classification of subgroups by geography, and consider patterns of
student placement as the metric by which states are grouped. Furthermore, instead of considering each placement category separately, as they would be in a variable-centered approach, cluster analysis allows for examination of nuanced differences in levels of all three-placement categories.

Results

States varied in their placement patterns into Categories A-C, as well as the percentages of all IDEA-eligible students within the state, as seen in Table 1.

Examination of a dendrogram associated with the full sample of states and the District of Columbia \((n = 51)\) revealed four distinct clusters of students within three educational placement categories (Highly inclusive, Moderately inclusive, Moderately restrictive, and Highly restrictive). As seen in Table 2, one-way ANOVAs confirmed that the four clusters varied significantly in their percentages of students in each placement category.

The breakdown of state by cluster is displayed in Table 3. The first cluster \((n = 13)\) was labeled as highly inclusive because this cluster had the highest average percentage of students in Category A, and lowest levels Category B and Category C settings. The second cluster \((n = 12)\) had a low percentage of students in the Categories B and C, so it was labeled moderately inclusive. The third cluster \((n = 15)\) was labeled as moderately restrictive because it contained a high percentage of students in the Categories B and C relative to the other clusters. The fourth and final cluster \((n = 11)\) was labeled as highly restrictive because the states in this cluster had a significantly higher percentage of students in separate school placements (Category C) compared with the other three clusters. Additionally, this cluster had the lowest percentage of students in Category A settings.

We then examined if the clusters varied by state-level characteristics. Table 4 displays only those covariates that significantly differed among the clusters. The highly restrictive cluster included states that were more urban, were more densely populated, had a higher percentage of Black citizens, lower graduation rates, and more students receiving free or reduced lunch than the other three clusters of states. Furthermore, the highly restrictive cluster consisted of states with citizens with higher median income and higher per pupil spending than the other clusters. In a sense, the highly restrictive cluster represents the most minoritized students (Black students, urban, and receiving free or reduced lunch) and some of the most privileged citizens (highest income and highest per pupil spending). The moderately restrictive cluster included states with higher poverty, lower median income, lower per pupil spending, lower population density, and lower percentages of citizens with high school diplomas than states in the other clusters. The highly inclusive cluster includes states that are more rural, lowest in poverty, had highest graduate rates, highest proportion of citizens having high school diplomas, the greatest percentage of White citizens. This cluster, in many respects, represents the most privileged citizens in terms of wealth and Whiteness. There were no variables in which the moderately inclusive cluster represented the most or least degree.

It is further important to note that additional variables, including disability density within a state (represented by the percentage of students with IEPs), proportion of a state’s Hispanic population, and the proportion of citizens in a state with bachelor’s degrees were found to have no statistical significance in this analysis.

Discussion

Limitations

Several limitations impact the interpretation of these results. First, the data collection systems in place to collect economic and demographic variables through the IDEA Data Accountability System may contain measurement errors that can impact the accuracy of data reporting. Thus, the accuracy of data from local and state education agencies, including percentages of time in various settings, warrants further investigation. Similarly, the diagnostic and administrative labels of autism spectrum disorders may result in variability in state definitions of ASD, and therefore the number of students with ASD educated in each state. Similarly, because ASD exists along a continuum of support needs, it is uncertain how different states categorize and support stu-
## TABLE 1

Percent of Students in with an ASD in Total and in Category A-C Placement by State (N = 51)

<table>
<thead>
<tr>
<th>State</th>
<th>Percent of all IDEA-eligible students with an ASD</th>
<th>Category A Placement Percent</th>
<th>Category B Placement Percent</th>
<th>Category C Placement Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>AL</td>
<td>6.8</td>
<td>59.6</td>
<td>22.5</td>
<td>4.2</td>
</tr>
<tr>
<td>AK</td>
<td>5.8</td>
<td>34.6</td>
<td>21.2</td>
<td>7.2</td>
</tr>
<tr>
<td>AZ</td>
<td>7.7</td>
<td>36.3</td>
<td>16.9</td>
<td>5.3</td>
</tr>
<tr>
<td>AR</td>
<td>6.2</td>
<td>33.4</td>
<td>39.1</td>
<td>2.7</td>
</tr>
<tr>
<td>CA</td>
<td>10.4</td>
<td>33.3</td>
<td>42.0</td>
<td>8.5</td>
</tr>
<tr>
<td>CO</td>
<td>5.9</td>
<td>53.7</td>
<td>19.3</td>
<td>4.4</td>
</tr>
<tr>
<td>DC</td>
<td>5.1</td>
<td>25.2</td>
<td>36.3</td>
<td>15.7</td>
</tr>
<tr>
<td>CT</td>
<td>10.6</td>
<td>51.7</td>
<td>11.9</td>
<td>18.7</td>
</tr>
<tr>
<td>DE</td>
<td>6.0</td>
<td>28.0</td>
<td>39.4</td>
<td>27.1</td>
</tr>
<tr>
<td>FL</td>
<td>6.9</td>
<td>34.2</td>
<td>46.0</td>
<td>9.8</td>
</tr>
<tr>
<td>GA</td>
<td>7.6</td>
<td>40.7</td>
<td>36.9</td>
<td>3.2</td>
</tr>
<tr>
<td>HI</td>
<td>7.0</td>
<td>30.5</td>
<td>35.7</td>
<td>1.3</td>
</tr>
<tr>
<td>ID</td>
<td>8.8</td>
<td>48.0</td>
<td>28.5</td>
<td>1.1</td>
</tr>
<tr>
<td>IL</td>
<td>6.8</td>
<td>32.8</td>
<td>30.6</td>
<td>15.2</td>
</tr>
<tr>
<td>IN</td>
<td>8.4</td>
<td>53.7</td>
<td>25.9</td>
<td>5.8</td>
</tr>
<tr>
<td>IA</td>
<td>1.1</td>
<td>64.5</td>
<td>8.2</td>
<td>2.6</td>
</tr>
<tr>
<td>KS</td>
<td>5.2</td>
<td>43.9</td>
<td>25.8</td>
<td>4.3</td>
</tr>
<tr>
<td>KY</td>
<td>5.4</td>
<td>41.9</td>
<td>29.8</td>
<td>1.9</td>
</tr>
<tr>
<td>LA</td>
<td>5.3</td>
<td>27.9</td>
<td>46.4</td>
<td>2.8</td>
</tr>
<tr>
<td>ME</td>
<td>9.0</td>
<td>41.8</td>
<td>25.5</td>
<td>6.8</td>
</tr>
<tr>
<td>MD</td>
<td>9.9</td>
<td>42.0</td>
<td>27.3</td>
<td>17.0</td>
</tr>
<tr>
<td>MA</td>
<td>8.2</td>
<td>37.9</td>
<td>30.3</td>
<td>16.5</td>
</tr>
<tr>
<td>MI</td>
<td>8.1</td>
<td>46.0</td>
<td>25.1</td>
<td>13.1</td>
</tr>
<tr>
<td>MN</td>
<td>13.3</td>
<td>53.7</td>
<td>18.7</td>
<td>5.6</td>
</tr>
<tr>
<td>MS</td>
<td>5.5</td>
<td>43.5</td>
<td>36.9</td>
<td>3.5</td>
</tr>
<tr>
<td>MO</td>
<td>7.7</td>
<td>35.2</td>
<td>25.3</td>
<td>8.3</td>
</tr>
<tr>
<td>MT</td>
<td>3.2</td>
<td>37.3</td>
<td>29.6</td>
<td>1.1</td>
</tr>
<tr>
<td>NE</td>
<td>5.8</td>
<td>62.3</td>
<td>16.8</td>
<td>5.3</td>
</tr>
<tr>
<td>NV</td>
<td>9.0</td>
<td>41.8</td>
<td>30.9</td>
<td>1.3</td>
</tr>
<tr>
<td>NH</td>
<td>7.4</td>
<td>32.9</td>
<td>20.0</td>
<td>8.5</td>
</tr>
<tr>
<td>NJ</td>
<td>6.8</td>
<td>23.9</td>
<td>34.1</td>
<td>27.3</td>
</tr>
<tr>
<td>NM</td>
<td>4.3</td>
<td>29.1</td>
<td>49.1</td>
<td>1.1</td>
</tr>
<tr>
<td>NY</td>
<td>6.3</td>
<td>25.3</td>
<td>42.5</td>
<td>22.2</td>
</tr>
<tr>
<td>NC</td>
<td>7.4</td>
<td>39.9</td>
<td>39.5</td>
<td>3.4</td>
</tr>
<tr>
<td>ND</td>
<td>6.2</td>
<td>58.1</td>
<td>14.4</td>
<td>4.7</td>
</tr>
<tr>
<td>OH</td>
<td>7.4</td>
<td>42.6</td>
<td>26.3</td>
<td>11.2</td>
</tr>
<tr>
<td>OK</td>
<td>4.2</td>
<td>41.7</td>
<td>31.6</td>
<td>0.6</td>
</tr>
<tr>
<td>OR</td>
<td>10.9</td>
<td>50.9</td>
<td>29.5</td>
<td>2.8</td>
</tr>
<tr>
<td>PA</td>
<td>8.4</td>
<td>43.5</td>
<td>22.8</td>
<td>10.7</td>
</tr>
<tr>
<td>RI</td>
<td>9.0</td>
<td>49.5</td>
<td>23.2</td>
<td>14.3</td>
</tr>
<tr>
<td>SC</td>
<td>5.0</td>
<td>29.8</td>
<td>49.8</td>
<td>2.7</td>
</tr>
<tr>
<td>SD</td>
<td>4.9</td>
<td>33.9</td>
<td>22.2</td>
<td>9.6</td>
</tr>
<tr>
<td>TN</td>
<td>5.7</td>
<td>42.7</td>
<td>34.0</td>
<td>3.0</td>
</tr>
<tr>
<td>TX</td>
<td>9.1</td>
<td>43.8</td>
<td>36.1</td>
<td>1.5</td>
</tr>
<tr>
<td>UT</td>
<td>6.6</td>
<td>34.3</td>
<td>33.8</td>
<td>6.4</td>
</tr>
<tr>
<td>VT</td>
<td>7.4</td>
<td>54.3</td>
<td>15.8</td>
<td>7.7</td>
</tr>
<tr>
<td>VA</td>
<td>9.2</td>
<td>40.0</td>
<td>31.3</td>
<td>7.3</td>
</tr>
<tr>
<td>WA</td>
<td>8.0</td>
<td>34.4</td>
<td>37.5</td>
<td>1.5</td>
</tr>
<tr>
<td>WV</td>
<td>3.9</td>
<td>33.7</td>
<td>37.5</td>
<td>1.0</td>
</tr>
<tr>
<td>WI</td>
<td>8.2</td>
<td>52.0</td>
<td>19.8</td>
<td>2.8</td>
</tr>
<tr>
<td>WY</td>
<td>6.0</td>
<td>46.5</td>
<td>19.9</td>
<td>2.4</td>
</tr>
<tr>
<td>U.S. Mean</td>
<td>7.7</td>
<td>40.9</td>
<td>29.5</td>
<td>8.9</td>
</tr>
</tbody>
</table>

Source: www.ideadata.org.
dents along this spectrum. Further research is needed to describe this variability and the impact of these sources of measurement error on reported data.

State Placement Patterns

This analysis reveals that a variety of economic, demographic, and educational factors are associated with educational placement of students with ASD. This analysis found students with ASD residing in states that are more rural, have more adults with high school diplomas, more White citizens, and higher graduation rates are more likely to be educated in inclusive settings. Students with ASD residing in states that are more urban, have a higher population density, more Black citizens, more students receiving a free or reduced cost lunch, higher median income, and higher per pupil spending are more likely to be educated in the most restrictive settings.

Inclusive education has often been associated with more economically and racially privileged groups. Specifically, children from higher socioeconomic (SES) backgrounds are more likely to receive less restrictive placements than children from lower SES backgrounds (Szumski & Karwowski, 2012). Typically, families must advocate for less restrictive

### TABLE 2

<table>
<thead>
<tr>
<th>Placement Categories</th>
<th>Clusters</th>
<th>F</th>
<th>Tukey HSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category A 80% or more</td>
<td>C1: Highly Inclusive n = 13</td>
<td>55.20</td>
<td>25.58</td>
</tr>
<tr>
<td></td>
<td>C2: Moderately Inclusive n = 12</td>
<td>42.47</td>
<td>35.76</td>
</tr>
<tr>
<td></td>
<td>C3: Moderately Restrictive n = 15</td>
<td>25.58</td>
<td>35.76</td>
</tr>
<tr>
<td></td>
<td>C4: Highly Restrictive n = 11</td>
<td>25.58</td>
<td>35.76</td>
</tr>
</tbody>
</table>

Note. * p < .001.

### TABLE 3

<table>
<thead>
<tr>
<th>States in each Cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highly Inclusive (n = 13)</td>
</tr>
<tr>
<td>Alabama, Idaho, Alaska, Arizona, California</td>
</tr>
<tr>
<td>Colorado, Kansas, Arkansas, California, D.C.</td>
</tr>
<tr>
<td>Connecticut, Kentucky, Georgia, Hawaii</td>
</tr>
<tr>
<td>Indiana, Massachusetts, Hawaii, Illinois, Iowa</td>
</tr>
<tr>
<td>Michigan, Missouri, Montana, Ohio</td>
</tr>
<tr>
<td>Minnesota, Montana, Maine, Maryland</td>
</tr>
<tr>
<td>North Dakota, Ohio, Idaho, Oklahoma, Oklahoma</td>
</tr>
<tr>
<td>Nebraska, Oklahoma, Mississippi, North Carolina, Nevada</td>
</tr>
<tr>
<td>New Hampshire, Pennsylvania, North Carolina, Nevada</td>
</tr>
<tr>
<td>Oregon, South Dakota, Virginia, Utah, Washington, West Virginia</td>
</tr>
</tbody>
</table>

placements for their children, but often families from lower SES backgrounds lack the resources for this type of sustained advocacy (Wakelin, 2008). Additionally, African-American, Hispanic, Native American, and English Language Learners have a higher chance of being placed in more restrictive placements than White students (de Valenzuela, Copeland, Huaqing Qi, & Park, 2006; Misra, 2006). Lastly, students with high-incidence disabilities (i.e., learning disabilities, speech/language impairments) are more likely to be placed in less restrictive settings than students with more significant disabilities (i.e., autism, intellectual disabilities, multiple disabilities; Misra, 2006). This analysis confirms these findings, indicating that states with citizens who are wealthier and Whiter are more likely to provide inclusive services to students with ASD.

However, the present analysis found highly restrictive states also included indicators of privilege (higher median income and higher per pupil spending) than states in other clusters. The highly restrictive states also had in common a number of indicators of low-privilege, including population density, higher percentage of Black citizens, and the percentage of students receiving a free or low cost lunch. These findings suggest highly restrictive placements may occur due to family choice in some instances (i.e., privileged families selecting highly restrictive settings due to assumptions about the effectiveness of those settings), while less privileged families may have no other options. Lauderdale-Littin, Howell, and Blacher (2013) similarly report that students with ASD from higher incomes were more likely to be educated in more restrictive settings. The benefits of highly restrictive placements reflect a set of assumptions about the unique opportunities of these settings, including access to distraction-free environments, specialized curriculum/instruction, behavioral supports, and development of a community of learners which are, in fact, rarely realized in these settings (Causton-Theoharis, Theoharis, Orsait, & Cosier, 2011). However, the assumptions about beneficial outcomes associated with highly restrictive settings for students with the greatest learning and support needs persist, which may influence privileged families to seek these placements. Conversely, families who have less priv-

<table>
<thead>
<tr>
<th>TABLE 4</th>
<th>Clusters by State Demographic, Student, and Disability Density Characteristics (N = 51)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariate</td>
<td>M within Cluster</td>
</tr>
<tr>
<td></td>
<td>C1: Highly Inclusive</td>
</tr>
<tr>
<td>Location and population</td>
<td></td>
</tr>
<tr>
<td>% Rural population</td>
<td>31.87</td>
</tr>
<tr>
<td>% Urban population</td>
<td>68.13</td>
</tr>
<tr>
<td>Persons/sq. mile</td>
<td>148.20</td>
</tr>
<tr>
<td>Education rates</td>
<td></td>
</tr>
<tr>
<td>HS graduates age 25+</td>
<td>89.42</td>
</tr>
<tr>
<td>Ethnicity within state</td>
<td></td>
</tr>
<tr>
<td>White (not Hispanic)</td>
<td>80.81</td>
</tr>
<tr>
<td>Black</td>
<td>7.22</td>
</tr>
<tr>
<td>SES indicators</td>
<td></td>
</tr>
<tr>
<td>% Free/reduced lunch</td>
<td>40.21</td>
</tr>
<tr>
<td>Median income</td>
<td>55316.50</td>
</tr>
<tr>
<td>% below poverty</td>
<td>12.39</td>
</tr>
<tr>
<td>Educational characteristics</td>
<td></td>
</tr>
<tr>
<td>Graduation rates (2011)</td>
<td>83.31</td>
</tr>
<tr>
<td>Per pupil spending</td>
<td>11752.80</td>
</tr>
</tbody>
</table>

Note. ***p < .001, **p < .01, *p < .05, †p = .062.
ilege may have less opportunity to seek and obtain any placement other than what is directly offered them by the local school district, which may result in a disproportionate number of students of color and lower SES backgrounds being placed in the most restrictive settings, despite any wishes of their families.

Lastly, the disproportionate identification of students with ASD from various ethnic backgrounds may impact placement rates. Black and Hispanic students continue to be under-identified for administrative prevalence of ASD compared to White students (Travers et al., 2014). Failure to identify students with ASD who are of color, while simultaneously placing more students of color in the most restrictive settings, may impact placement rates across states while perpetuating the Whiteness of inclusive settings.

**Implications**

Researchers have asserted that placement in less-restrictive settings conceptually (Jackson, Ryndak, & Wehmeyer, 2009) and practically (e.g., Kurth & Mastergeorge, 2012) benefit students with disabilities, including students with ASD. The present analysis, however, found that placement in less-restrictive settings varied along a number of economic, demographic, and educational variables, suggesting an inequitable access to the LRE for students with ASD, suggesting a need for further research into the factors that contribute to this outcome.

The present analysis focused on placement of students with ASD for the year 2012, including analysis of Census and IDEA data. However, these data provide simply a snapshot in time. Further analyses of data over the past decade may reveal trends in placement patterns over time, particularly following the implementation of IDEA 2004 and No Child Left Behind Act of 2001, both of which strengthened federal commitments to access and progress in general education (e.g., Wilson, Kim, & Michaels, 2013). Similarly, it is possible that trends in placement may correlate with trends in prevalence of ASD. Specifically, the Centers for Disease Control report that ASD prevalence increased from 1 in 150 in the year 2000, to 1 in 68 in the year 2010; it is unknown to what extent changes in placement patterns correspond with increasing prevalence over this same time period. It is possible that as schools have grappled with the issue of serving more students with ASD, they made changes in placement patterns as the increasing numbers of students impacted existing placements. Again, a fuller picture of placement patterns over time may indicate the extent to which progress is, or is not, being made in gaining access to the LRE for all students with ASD.

Furthermore, existing research has documented disproportionate identification of students with ASD. Specifically, White and Asian/Pacific Islander students tend to be over-represented in the ASD category, whereas Black, Hispanic, and American Indian/Alaska Native students tend to be under-represented (Marks & Kurth, 2014; Travers et al., 2014). Marks and Kurth further found that states with a higher ASD prevalence rate demonstrated less disproportional identification of students with ASD by race than lower prevalence states, suggesting states with higher prevalence rates may have systems and structures in place to develop statewide efforts related to ASD awareness which may impact disproportional identification. The findings of this analysis indicate minoritized students with ASD are more likely to be placed in more restrictive settings. Further understanding and development of policies, including funding mechanisms and state placement guidelines, that support less-restrictive placement patterns, particularly for minoritized youth, are needed.

Finally, clarification regarding how IEP teams arrive at placement decisions is needed, including the role of biases and assumptions about students. For example, in a study of first-grade teacher opinions regarding educational placement, Segall and Campbell (2014) found teachers were more likely to place a hypothetical student described as having an intellectual disability and ASD in a more restrictive setting compared to a hypothetical student with average intelligence and ASD. On the other hand, Segall and Campbell also report teachers felt their own classrooms would be less appropriate for students with ASD than a hypothetical other classroom, although teachers who felt stakeholders such as parents and administrators favored inclusion and teachers with greater self-reported com-
petence were more likely to suggest a less restrictive placement. In a similar analysis of assumptions and biases, Begeer and colleagues (2009) found a referral bias in ASD, in that hypothetical students from ethnic minority groups were less likely to be referred for ASD identification than majority-group hypothetical students (Begeer, El Bouk, Boussaid, Terwogt, & Koot, 2009). Together, such research indicates the persistent impact of assumptions and biases on students with ASD, and the impact such biases may have on access to the LRE. The present analysis found a bifurcation in the highly segregated states, so that both high-privilege and low-privilege students were placed in restrictive settings. The biases and assumptions that underpin these findings need further exploration.

References


Received: 5 February 2015
Initial Acceptance: 16 April 2015
Final Acceptance: 28 May 2015
Meta-Analysis of Pivotal Response Training for Children with Autism Spectrum Disorder

Gulden Bozkus-Genc and Serife Yucesoy-Ozkan
Anadolu University

Abstract: The main purpose of this study was to review pivotal response training and examine the efficacy of pivotal response training for children with autism spectrum disorder. The other purposes of study were to (a) examine the characteristics of participants and components of the intervention in which pivotal response training was used; (b) determine the level of efficacy of pivotal response training to teach various behaviors to children with autism spectrum disorder; (c) determine whether the effectiveness of pivotal response training differed in terms of characteristics of the intervention; and (d) determine whether percentage of nonoverlapping data, percentage of nonoverlapping corrected data, and percentage of data points exceeding median were correlated. In this study, 34 single case research articles conducted with individuals with autism spectrum disorder and published in a peer-reviewed journal in between 1979–2012 were examined. Articles primarily were descriptively analyzed and then examined by use of meta-analysis. According to results, in half of the studies, treatment integrity was assessed, generalization and maintenance data were collected; in only a quarter of the studies, social validity data were collected. Pivotal response training that focused on two of the three core features of autism spectrum disorder were found effective in influencing individual outcomes. Results also indicated that percentage of nonoverlapping data, percentage of nonoverlapping corrected data, and percentage of data points exceeding median were correlated.

One of the evidence-based interventions used to teach individuals with autism spectrum disorder (ASD) is pivotal response training (PRT). PRT (also referred to as pivotal response teaching, pivotal response treatment, pivotal response therapy, or pivotal response intervention in the literature) is a form of naturalistic behavioral intervention based on the principles of applied behavior analysis, which assume that children’s impairments can be improved with environmental manipulations such as reinforcement, consequences, and extinction (Koegel, Koegel, & Carter, 1999; Stahmer, Suhrheinrich, Reed, Bolduc, & Schreibman, 2010). PRT was developed to facilitate generalization, increase spontaneity, reduce prompt dependency, and increase motivation (Suhrheinrich, 2010). Specific components of PRT include providing clear and appropriate cues, allowing the child to choose an activity and make choices within an activity, turn-taking, interspersing maintenance tasks with acquisition tasks, reinforcing the child’s attempts, responding to multiple cues, and providing contingent reinforcement directly related to the child’s response (Koegel, Koegel, & McNerney, 2001).

The basic premise of PRT is that changes in certain pivotal areas of behavior will trigger changes in other behavioral areas (Koegel et al., 2001). These pivotal areas are motivation, responsivity to multiple cues, self-management, and self-initiations (Koegel & Koegel, 2006, 2012; Koegel, Koegel, Shoshan, & McNerney, 1999). According to PRT, these skills
are pivotal because they are the foundational skills enabling learners with ASD to make widespread and generalized improvements in many other areas (National Professional Development Center on Autism Spectrum Disorders, 2012).

Numerous research studies support the effectiveness of PRT. PRT improved several language functions, including speech imitation (Laski, Charlop, & Schreibman, 1988), labeling (Koegel, Camarata, Valdez-Menchaca, & Koegel, 1998), asking-questions (Koegel, Camarata, Koegel, Ben-Tall, & Smith, 1998), spontaneous speech (Laski et al., 1988), and conversational communication (Koegel et al., 1998). PRT has been adapted for use in teaching social skills including self-initiations (Koegel, Carter, & Koegel, 2003), joint attention (Whalen & Schreibman, 2003), sociodramatic play (Thorp, Stahmer, & Schreibman, 1995), peer social interaction (Pierce & Schreibman, 2010), and academic skills (Koegel, Singh, & Koegel, 2010). PRT has also been used to decrease problem behaviors (Baker-Ericzen, Stahmer, & Burns, 2007).

Examining the PRT studies to date reveals that three reviews have been conducted, but no meta-analyses. The purpose of the first review (Masiello, 2007) was to determine the effectiveness of PRT to improve social-emotional and communicative behavioral outcomes of children with ASD. For this purpose, 13 research studies published between 1988 and 2003 were descriptively analyzed. It was reported that PRT was effective in improving the social-emotional and communicative behavioral outcomes of children with ASD. The second review (Bozkus-Genc & Vuran, 2013) was a qualitative document analysis of the different variables in 23 studies published from 1987 to 2011 in which PRT was used to teach social skills to children with ASD. Once again, PRT was reported to be effective for teaching social skills to children with ASD. In the third review (Toper-Korkmaz & Diken, 2013), 16 studies published from 1995 to 2011 using PRT were examined descriptively. PRT was indicated as an effective method to teach expressive language skills, social skills, and play skills. Although PRT was suggested as an effective method in the aforementioned reviews, quantitative data regarding the significance of the functional relationship, which might show the effectiveness of PRT, were not reported. Effect size calculations must be included to describe the effectiveness of an intervention in a clear and explicit way (Kazdin, 1982). Therefore, it is important to include effect size calculations in studies to examine the effectiveness of PRT in individuals with ASD.

Several computational methods have been proposed to calculate effect size in single case research (SCR) studies (Wolery, Busick, Reichow, & Barton, 2010). The most widely used methods to calculate effect sizes in SCR studies are either regression or non-regression-based (Olive & Franco, 2007; Olive & Smith, 2005; Wendt, 2009). Several studies have examined the relationship between regression and non-regression-based effect size calculations (e.g., Olive & Smith, 2005; Parker & Hagan-Burke, 2007; Wolery et al., 2010). However, researchers still debate which calculation method should be used. Percentage of nonoverlapping data (PND) is the oldest and most widely used effect size calculation method (Mastropieri & Scruggs, 1985–1986; Parker, Vannest, & Davis, 2011; Scruggs, Mastropieri, & Casto, 1987; Scruggs, Mastropieri, Cook, & Escobar, 1986); however, it does not have sufficient sensitivity (Faith, Allison, & Gormann, 1996). If there is a very high data point in the baseline level, the intervention may seem to be ineffective even though it is effective. Conversely, if there is even a very small increase in the baseline data, the intervention may appear to be effective even though it is ineffective (Faith et al., 1996). In order to eliminate these limitations of PND, percentage of nonoverlapping corrected data (PNCD) and percentage of data points exceeding the median (PEM) were developed. PNCD suggests a data correction process before calculating the PND in order to separate the possible trend from the data before the intervention (Manolov & Solanas, 2009). PEM is based on the assumption that the median best summarizes the data in the baseline level (Ma, 2006). Even though many studies have examined the PND and PEM calculations (Parker & Hagan-Burke, 2007; Wolery et al., 2010), only a limited number of them has been conducted on PNCD (Manolov & Solanas, 2009). Therefore, it is necessary to examine whether PNCD is consistent with other
methods based on overlap. The purpose of this study was to examine the efficacy of PRT. Thus, in the present study, it was sought answers to the following questions: (a) what are the characteristics of the participants and components of the intervention in which PRT was used? (b) what is the level of the effectiveness of PRT with regard to teaching various behaviors to children with ASD? (c) does the effectiveness of PRT differ in terms of the characteristics of the intervention? (d) are the PND, PNCD, and PEM effect size calculations correlated?

Method

Study Identification

A comprehensive literature review was conducted in order to identify the research studies that would be included in the meta-analysis. Five different methods were used to identify the studies: searching the related terms in the subject indexes, searching the thesis databases, footnote chasing, hand searching, and consulting with researchers. One method used to identify the studies was to search certain keywords in the subject indexes in online databases. The online databases that we searched are given in alphabetical order, as follows: Academic Search Complete, Cambridge Journals Online, Ebrary, Oxford Journals Online, Psychology and Behavioral Science, Sage Journal, Science Direct Journals, SocINDEX with Full Text, Springer LINK Contemporary, Taylor and Francis Journals, Wiley Black, and Wilson Select Plus. As articles for studies in education, psychology, and sociology most commonly appear in these databases, the abovementioned databases were preferred for the literature review. In addition, Google Scholar was used to search for studies. Keywords such as natural language paradigm (NLP), pivotal response teaching, pivotal response training, pivotal response treatment, pivotal response therapy, pivotal response intervention, and PRT and the combination of these terms with autism, initiation, self-initiation, joint attention, motivation, empathy, response to multiple cues, social skills, communication skills, and play skills were used.

The second method employed was to search the thesis databases. Proquest Dissertations and Theses database was reviewed to obtain information not available through the other sources. The same key words as detailed in the paragraph above were used to search the online databases. Dissertation abstracts, which matched these key words, were then reviewed manually to determine whether they met the inclusion criteria. The references sections of related theses were examined. The third method for locating studies was footnote chasing (or backward chaining), which included an examination of the references in articles of related research studies. The reference lists of books and articles of interest were examined; this led to the rapid identification of the primary studies. The fourth method used to identify studies was hand searching. This procedure enabled the search for articles that had just been published and, therefore, were not yet included in the online databases. So, hand searches examined additional studies. The fifth method used to identify articles involved consultations with researchers. After completing the search of the online databases, we requested the full text of the articles for which only summaries had been obtained from the authors or institutions (e.g., articles not found in the online databases, research reports not obtainable through other means). We emailed researchers and institutions to request papers, including research studies related to instructional practices in PRT and those conducted with children with ASD. In addition to the abovementioned methods, we reviewed the table entitled “Empirical Support for Pivotal Response Treatment” on the University of California Santa Barbara (UCSB) Koegel Autism Center web page. This table consists of the core research studies conducted from 1979 to 2010 on the pivotal area of motivation and initiations.

The literature review and the abovementioned methods yielded a pool of 69 studies. The summary of each of the studies was examined. Next, we scanned and skimmed the full text of each of the papers. Scanning and skimming revealed that four of the studies (6%) were qualitative, 11 (16%) were informative articles and reviews, and 54 (78%) were experimental studies. The experimental studies (54 research papers) were examined.
in detail based on the inclusion criteria listed in the following section.

Inclusion and Exclusion Criteria

The studies to be included in this meta-analysis met the following criteria: (a) the study was published in a peer-reviewed journal from 1979 to 2012; (b) the article was written in English; (c) the participants in the study were 1–13 years of age and diagnosed with ASD; (d) the effectiveness of PRT was examined in the study; (e) one of the SCR designs was used in the study; (f) the effect of the independent variable was shown by a line graph; and (g) there was at least one data point in the baseline phase and the intervention phase.

The inclusion criteria were chosen for specific reasons. First, since the first study on PRT was published in 1979, studies conducted since 1979 were included. Second, articles written in English were included because English is the international scientific language, and it is easier to access articles written in English. Third, almost all of the studies examining the effectiveness of PRT were conducted with children with ASD in the range of early childhood to the middle school years; therefore, we included studies of children with ASD from 1 to 13 years old. Fourth, most of the studies employed a SCR design. Because the inclusion of group experimental designs with SCR designs in meta-analysis studies might generate inaccurate results. Thus, studies only conducted with SCR design were included in this study. Line graph is used in a widespread manner in the single subject designs. Because in the other graphs (e.g., bar, column) are only shown average phase values. So, these graphs do not give information especially in the event of leaps or fluctuations in the data points. For this reason, studies using line graph were included. Last, in order to calculate effect sizes and to examine the functional relationship between the baseline and intervention phases, the articles that reported at least one data point in both the baseline and intervention phases were included.

Twenty studies were excluded from this meta-analysis. Two articles that included teaching families or paraprofessionals to use PRT and examining the degree of their use of PRT were excluded. Ten articles based on group experimental designs were excluded. One article was excluded in which the AB design did not allow for the formation of a functional relationship between the independent and the dependent variable. Two studies that used an ABC design and comparison-based alternating treatment designs were excluded. One article that did not include data points in the baseline phase was excluded. Four articles were excluded in which the effects of the independent variable were demonstrated in column chart rather than a line graph. A final total of 34 research studies published in 12 different journals were included in the meta-analysis. The list of research studies included in this meta-analysis can be obtained from the authors.

Coding Procedure

In coding procedure, first, the file of articles was created, all the articles were carefully read and summarized in a Microsoft Office Word® file. Then, a coding key and manual that explained how to code the studies were prepared by the authors following the coding systems used by Horner et al. (2005) and Odom et al. (2010). Taking into account this coding key and manual, each article was coded into a Microsoft Office Excel® file by the first author. Studies were coded in the following terms: (a) the study tag (article number, author name, year published, and the journal name); (b) participant characteristics (diagnosis, mean age, and gender); and (c) intervention characteristics (settings, instructional modification, intervener, research model, dependent variable, inter-observer reliability, treatment integrity, progress, follow-up, generalization, and social validity).

Key considerations were taken into account while the articles were coded. First, the mean age of the participants was calculated by converting the age into months (e.g., 6 years 4 months = 76 months), summing the ages of all of the participants, and dividing the sum by the number of participants; this mean value was coded into the coding key. Participant gender was coded as described in the article. However, if the gender of the participants was unclear, the entire article was scanned and statements of gender were taken into account when gender was coded. The study setting was
coded as clinic, school, home, community setting, or unspecified. The intervener was coded as expert, family, peer, or other. The research models were coded as multiple baseline designs, ABA design, or ABAB design. The dependent variables were coded as communication and interaction skills, social skills, play skills, academic skills, or multiple skills. Progress, follow-up, generalization, inter-observer reliability, treatment integrity, and social validity were coded as either available or not available.

**Effect Size Calculations**

In order to indicate the effectiveness of the SCR designs, the baseline and the intervention phases were compared by visual analysis, and effect size was calculated. Visual analysis involves reaching a conclusion about the effects of various independent variables on dependent variables by visually examining the graphed data. Meaningful changes in the dependent variable should be apparent when displayed graphically, and emphasis is placed on the believability of the observed change behavior (Poane, Rihgdahl, Kelley, & Glover, 2011). Effect sizes were calculated using the following three methods: PND (Scruggs et al., 1987), PNCD (Manolov & Solanas, 2009), and PEM (Ma, 2006). Because a variety of SCR designs were represented in this meta-analysis, different strategies were identified in order to calculate effect size metrics. The effect sizes for the various SCR designs were calculated as follows.

For the ABA design, effect size was calculated for the baseline and intervention phases for each of the participants (A1 and B1). For the ABAB design, effect size was calculated for the first baseline and first intervention and for the second baseline and second intervention phases for each of the participants (A1 and B1, A2 and B2). For the multiple baseline design, effect size was calculated for each of the behaviors; then, the individual scores were averaged to obtain the average score for the study.

In order to calculate the PND score, a line parallel to the horizontal axis is drawn from the highest data point in the baseline throughout the intervention phase. Meaningful changes in the dependent variable should be apparent when displayed graphically, and emphasis is placed on the believability of the observed change behavior (Poane, Rihgdahl, Kelley, & Glover, 2011). Effect sizes were calculated using the following three methods: PND (Scruggs et al., 1987), PNCD (Manolov & Solanas, 2009), and PEM (Ma, 2006). Because a variety of SCR designs were represented in this meta-analysis, different strategies were identified in order to calculate effect size metrics. The effect sizes for the various SCR designs were calculated as follows.

For the ABA design, effect size was calculated for the baseline and intervention phases for each of the participants (A1 and B1). For the ABAB design, effect size was calculated for the first baseline and first intervention and for the second baseline and second intervention phases for each of the participants (A1 and B1, A2 and B2). For the multiple baseline design, effect size was calculated for each of the behaviors; then, the individual scores were averaged to obtain the average score for the study.

In order to calculate the PND score, a line parallel to the horizontal axis is drawn from the highest data point in the baseline throughout the intervention phase. For appropriate and inappropriate behaviors, the number of data points above and below this line, respectively, are divided by the total number of data points in the intervention phase, and then multiplied by 100 (Scruggs & Mastropieri, 1998). PND scores range from 0 to 100, and can be interpreted using the conventions established by Scruggs et al. (1986). PND scores >90, 70–79, 50–69, and <50 represent highly effective, fairly effective, questionable, and ineffective treatments, respectively (Scruggs & Mastropieri, 1998, 2001). In order to calculate PND scores, the graphs of each of the studies were digitally saved, then enlarged and printed. In total, 295 PND scores were calculated.

PNCD is a data correction procedure to be implemented prior to calculating the PND. The main aim of the PNCD is to eliminate a possible pre-existing data trend unrelated to the introduction of the intervention (Manolov & Solanas, 2009). In order to calculate PNCD scores, a difference series is obtained by subtracting the previous data point from every data point at the baseline level (nA-1). The mean of these newly computed values are calculated. Having calculated the mean, the trend correction factor for each of the data points is computed by multiplying the mean of the difference series by the order of the data point. After calculating the correction factors, a data correction process is applied that includes subtracting these correction factor scores from the original data points. After this correction is performed, a PND is calculated using the new data points. No score system exists to interpret the PNCD scores. Therefore, a PND score system was used to interpret the PNCD scores. Since electronic copies of 32 of the 34 articles were obtained before calculating a PNCD, hard copies of the articles were converted to soft copies by a digital scanner. The graphics of the soft copies of the articles were saved as pictures in JPEG format by using Adobe Acrobat 9 Pro software. Saved as pictures, the graphics were named according to their order of appearance in the article and given their final forms by editing them with Microsoft Office Picture Manager 2010 software. The second author performed this process. Using Plot Digitizer 2.5 for MacOS digitized the soft copy graphics. Plot Digitizer is a Java program used to digitize scanned plots of functional data (http://plotdigitizer.sourceforge.net). New graphics were drawn for
each of the participants and variables using the digitized data. These new graphics were visually compared with the original ones and found to be nearly identical in all cases. In total, 284 PNC scores were calculated using the digital data. For one of the articles, the graphics could not be viewed with Plot Digitizer. Therefore, its data points could not be digitized. Consequently, for that article, a PNC score was not calculated.

The PEM metric is computed by calculating the percentage of the treatment data points that do not overlap with the median baseline data point (Ma, 2006, 2009). In order to calculate a PEM, a horizontal line to the x-axis is drawn from the median of the data points in the baseline phase throughout the intervention phase. For appropriate and inappropriate behaviors, the number of data points above and below this line, respectively, is divided by the total number of data points in the intervention phase, resulting in the PEM score. PEM scores range from 0 to 1. PEM scores .90–1.0, .70–.89, and <.70 represent very effective, moderately effective, and questionable or ineffective treatments, respectively (Wendt, 2009; Wolery et al., 2010). PEM scores were multiplied by 100 in this study in order to obtain scores in percentages. To calculate PEM scores, the graphs of the studies were digitally saved, enlarged, and printed. In total, 295 PEM scores were calculated.

Correlation Analysis

IBM SPSS Statistics 21.0 for MacOS was used to compute Pearson’s correlation coefficients in order to examine the degree and direction of the relationship among the three different effect size calculation methods. To examine the individual effects of studies rather than using the average effect size for each study, every effect size calculated was used to calculate the correlation coefficient. A total of 284 effect sizes were used to calculate the correlation coefficient. Even though 295 effect sizes were calculated in this study, since the data points in two of the figures could not be digitized and PNC scores could not be calculated, PND and PEM scores for these figures were excluded. Thus, for each of the calculation methods, 284 effect size scores were used. We used a total of 852 effect size scores.

Reliability

In this meta-analysis, inter-rater reliability was assessed, both for the coding process and the effect size calculations, by the second author. Of all the studies, 14 (41%) were randomly selected for an inter-rater reliability assessment; the second author coded these studies. Next, the data coded by the first and the second authors were listed and every category was compared. Inter-rater reliability was obtained by a point-by-point ratio between the first author and the second author (Wolery, Bailey, & Sugai, 1988). There were 326 agreements and 34 disagreements. The inter-rater reliability was 95.1%.

To compute inter-rater reliability for the PND and PEM scores in all of the studies, the PNC scores of 11 of the studies (30%) were chosen. Since PNC decreases the probability of errors in mathematical calculations, only 30% of the research studies were selected to assess the inter-rater reliability of PNC. The calculations of the first author were compared with those of the second author. For PND calculations, there were 289 agreements (97.6%) and 6 disagreements. For the PEM calculations, there were 294 agreements (99.6%) and 1 disagreement. For the PNC calculations, there were 80 agreements (100%).

Results

Results for the participant and intervention characteristics are summarized in Table 1. As shown in Table 1, most of the studies (32.3%) were published from 2009 to 2012. The majority (44.1%) was published in journals related to ASD. As listed in Table 1, there were 125 participants in total across the 34 studies. 107 (85.6%) were diagnosed with ASD, eight (6.4%) were diagnosed with pervasive developmental disorder (PDD), and 10 (8%) had comorbid disabilities with ASD. The mean age of all of the participants was 4 years, 8 months (range 2 years, 5 months to 12 years, 8 months). In two studies (5.9%) the mean age of the participants was 1–3 years, in 21 studies (61.8%) it was 3–6 years, in eight studies (23.5%) it was 6–9 years, and in three studies (8.8%) it was >9 years. Moreover, 89 of the participants (71.2%) were male, 26 (20.8%) were female; for 10 studies (8%),
TABLE 1

Summary of Participant and Intervention Characteristics

<table>
<thead>
<tr>
<th>Participant Characteristics</th>
<th>f</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASD</td>
<td>107</td>
<td>85.6</td>
</tr>
<tr>
<td>ASD and additional disabilities</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>Pervasive Developmental Disorder</td>
<td>8</td>
<td>6.4</td>
</tr>
<tr>
<td>(n = 125)* (Σ 100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0–3 years</td>
<td>14</td>
<td>5.9</td>
</tr>
<tr>
<td>3–6 years</td>
<td>75</td>
<td>61.8</td>
</tr>
<tr>
<td>6–9 years</td>
<td>31</td>
<td>23.5</td>
</tr>
<tr>
<td>9 years and over</td>
<td>5</td>
<td>8.8</td>
</tr>
<tr>
<td>(n = 125)* (Σ 100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>89</td>
<td>71.2</td>
</tr>
<tr>
<td>Female</td>
<td>26</td>
<td>20.8</td>
</tr>
<tr>
<td>N/A</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>(n = 125)* (Σ 100)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Intervention Characteristics</th>
<th>f</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Setting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple</td>
<td>15</td>
<td>44.1</td>
</tr>
<tr>
<td>Clinic</td>
<td>9</td>
<td>26.4</td>
</tr>
<tr>
<td>School</td>
<td>5</td>
<td>14.7</td>
</tr>
<tr>
<td>Home</td>
<td>3</td>
<td>8.9</td>
</tr>
<tr>
<td>Community</td>
<td>2</td>
<td>5.9</td>
</tr>
<tr>
<td>(n = 34)** (Σ 100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructional Arrangement</td>
<td></td>
<td></td>
</tr>
<tr>
<td>One to one</td>
<td>26</td>
<td>76.5</td>
</tr>
<tr>
<td>Small group</td>
<td>3</td>
<td>8.8</td>
</tr>
<tr>
<td>N/A</td>
<td>5</td>
<td>14.7</td>
</tr>
<tr>
<td>(n = 34)** (Σ 100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intervener</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professional (researcher/clinician/therapist/teacher)</td>
<td>13</td>
<td>38.2</td>
</tr>
<tr>
<td>Parent/caregiver</td>
<td>8</td>
<td>23.5</td>
</tr>
<tr>
<td>Student (undergraduate/graduate student)</td>
<td>6</td>
<td>17.6</td>
</tr>
<tr>
<td>Peer</td>
<td>4</td>
<td>11.7</td>
</tr>
<tr>
<td>Multiple</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>(n = 34)** (Σ 100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Design</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple baseline</td>
<td>29</td>
<td>85.3</td>
</tr>
<tr>
<td>ABAB</td>
<td>4</td>
<td>11.8</td>
</tr>
<tr>
<td>ABA</td>
<td>1</td>
<td>2.9</td>
</tr>
<tr>
<td>(n = 34)** (Σ 100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dependent Variable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Language and communication/interaction skills</td>
<td>15</td>
<td>44.1</td>
</tr>
<tr>
<td>Multiple skills</td>
<td>11</td>
<td>32.5</td>
</tr>
<tr>
<td>Social skills</td>
<td>4</td>
<td>11.7</td>
</tr>
<tr>
<td>Play skills</td>
<td>3</td>
<td>8.8</td>
</tr>
<tr>
<td>Academic skills</td>
<td>1</td>
<td>2.9</td>
</tr>
<tr>
<td>(n = 34)** (Σ 100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reliability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inter-observer reliability</td>
<td>34</td>
<td>100</td>
</tr>
<tr>
<td>Treatment integrity</td>
<td>15</td>
<td>44.1</td>
</tr>
</tbody>
</table>

TABLE 1—(Continued)

<table>
<thead>
<tr>
<th>Intervention Characteristics</th>
<th>f</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maintenance, Generalization, and Social Validity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maintenance</td>
<td>15</td>
<td>44.1</td>
</tr>
<tr>
<td>Generalization</td>
<td>21</td>
<td>61.7</td>
</tr>
<tr>
<td>Social validity</td>
<td>8</td>
<td>23.5</td>
</tr>
</tbody>
</table>

f: Frequency.  
%: Percentage.  
n: Number of studies.  
* Total number of subjects.  
** Total number of research studies.  

information regarding gender was not included.

The settings were multiple in 15 of the studies (44.1%), clinic in nine (26.4%), school in five (14.7%), home in three (8.8%), and community settings in two (5.8%). PRT was presented through one-to-one instruction in 26 of the studies (76.5%), and through small group instruction in three of the studies (8.8%). In five of the studies (14.7%), no information was provided regarding the instructional arrangement. PRT was presented by professionals (clinicians, therapists, or teachers) in 13 of the studies (38.2%), by families/caregivers in eight (23.5%), by students (undergraduate or graduate) in six (17.6%), by peers in four (11.7%), and by more than one intervener in three (9%).

A multiple baseline design was used in 29 of the studies (85.3%), an ABAB design was used in four (11.8%), and an ABA design was used in one (2.9%). In 15 of the studies (44.1%), the dependent variables were communication and interaction skills. In 11 of the studies (32.5%) multiple skills were targeted. In four studies (11.7%) social skills, in three studies (8.8%) play skills, in one study (2.9%) academic skills were the dependent variables. In all of the studies (100%), PRT improved the dependent variables. In all of the studies (100%), inter-observer reliability data were gathered (range = 80–99). Treatment integrity was assessed in only 15 (44.1%) of the studies (range = 78.3–100) were assessed. Moreover, in 20 (55.8%), 15 (44.1%), and eight (23.5%) of the studies, generalization data, maintenance data, and social validity data were collected, respectively.
The effect sizes of the studies included in this meta-analysis are listed in Table 2. The table demonstrates that, for all of the studies, the mean PND score was 76.10% (range = 0–100%, standard deviation [SD] = 33.65), the mean PNCD score was 78.03% (range = 0–100%, SD = 34.38), and the mean PEM score was 89.34% (range = 0–100%, SD = 22.18). The level of effect for the PND, the PNCD, and the PEM scores of the studies included in this meta-analysis are listed in Table 3. As demonstrated in Table 3, with regard to PND scores, the PRT effect sizes were greater than 90% in 13 studies (38.2%), 70–89% in 11 studies, and below 70% in 10 studies (29.4%). With regard to PNCD scores, effect sizes were greater than 90% in 14 studies (41.1%), 70–89% in nine studies (26.4%), and below 70% in 10 studies (26.4%). With regard to PEM scores, effect sizes were greater than 90% in 27 studies (79.4%), 70–89% in four (11.7%), and below 70% in three studies (8.8%). The effect size scores of every study and each calculation type can be obtained from the authors.

The PND, PNCD, and PEM scores in this study were examined in terms of different variables. The results obtained are given in Table 4. As shown in Table 4, effect size scores were greater than 70% in all of the research models. PND scores for all of the dependent variables except play skills and social skills were greater than 70%. PNCD scores for all of the dependent variables except play skills were greater than 70%. PEM scores for all of the dependent variables except play skills were greater than 70%.

PEM scores for all of the dependent variables were greater than 70%. For all of the studies, regardless of whether inter-observer reliability, treatment integrity, maintenance, or generalization data were gathered, all three-effect size scores were greater than 70%. In the studies in which social validity data were gathered, PND and PNCD scores were below 70% whereas PEM scores were above 90%. In studies in which researchers did not collect social validity data, all three-effect size scores were greater than 70%.

In order to examine the degree and direction of the relationship among PND, PNCD, and PEM scores, correlation coefficients were computed; they are listed in Table 5. As shown in Table 5 there were strong, positive correlations between PND and PNCD (r = .749, p < .001), PND and PEM (r = .598, p < .001), and

### TABLE 2

<table>
<thead>
<tr>
<th>Calculation</th>
<th>N</th>
<th>Min.</th>
<th>Max.</th>
<th>X</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>PND</td>
<td>284</td>
<td>0.00</td>
<td>100</td>
<td>76.10</td>
<td>33.65</td>
</tr>
<tr>
<td>PNCD</td>
<td>284</td>
<td>0.00</td>
<td>100</td>
<td>78.03</td>
<td>34.38</td>
</tr>
<tr>
<td>PEM</td>
<td>284</td>
<td>0.00</td>
<td>100</td>
<td>89.34</td>
<td>22.18</td>
</tr>
</tbody>
</table>

N: Number of studies.  
Min.: Minimum value.  
Max.: Maximum value.  
X: Mean.  
SD: Standard deviation.

### TABLE 3

<table>
<thead>
<tr>
<th>Overlap Method</th>
<th>Highly Effective (&gt; 90%)</th>
<th>Fairly Effective (70-89%)</th>
<th>Questionable or Ineffective (&lt; 70%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>f</td>
<td>%</td>
<td>f</td>
</tr>
<tr>
<td>PND</td>
<td>13</td>
<td>38.2</td>
<td>11</td>
</tr>
<tr>
<td>PNCD</td>
<td>14</td>
<td>41.1</td>
<td>9</td>
</tr>
<tr>
<td>PEM</td>
<td>27</td>
<td>79.4</td>
<td>4</td>
</tr>
</tbody>
</table>

f: Frequency.  
%: Percentage.
TABLE 4

Means and standard deviations of the PND, PNCD, and PEM scores for the intervention characteristics

<table>
<thead>
<tr>
<th>Variables</th>
<th>PND</th>
<th>PNCD</th>
<th>PEM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>X</td>
<td>SD</td>
</tr>
<tr>
<td>Design</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple baseline</td>
<td>29</td>
<td>76.60</td>
<td>20.27</td>
</tr>
<tr>
<td>ABAB</td>
<td>4</td>
<td>91.22</td>
<td>15.20</td>
</tr>
<tr>
<td>ABA</td>
<td>1</td>
<td>97.2</td>
<td>0</td>
</tr>
<tr>
<td>Dependent variable</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Communication and interaction</td>
<td>13</td>
<td>90.27</td>
<td>19.25</td>
</tr>
<tr>
<td>skills</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple skills</td>
<td>12</td>
<td>76.18</td>
<td>16.96</td>
</tr>
<tr>
<td>Social skills</td>
<td>5</td>
<td>68.48</td>
<td>11.96</td>
</tr>
<tr>
<td>Play skills</td>
<td>3</td>
<td>57.3</td>
<td>20</td>
</tr>
<tr>
<td>Academic skills</td>
<td>1</td>
<td>84.2</td>
<td>0</td>
</tr>
<tr>
<td>Improvement</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>34</td>
<td>76.10</td>
<td>20.54</td>
</tr>
<tr>
<td>No</td>
<td>0</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Interobserver reliability</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>34</td>
<td>76.10</td>
<td>20.54</td>
</tr>
<tr>
<td>No</td>
<td>0</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Treatment integrity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>15</td>
<td>72.41</td>
<td>22.6</td>
</tr>
<tr>
<td>No</td>
<td>19</td>
<td>84.21</td>
<td>16.37</td>
</tr>
<tr>
<td>Maintenance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>15</td>
<td>76.3</td>
<td>19.05</td>
</tr>
<tr>
<td>No</td>
<td>19</td>
<td>81.14</td>
<td>20.88</td>
</tr>
<tr>
<td>Generalization</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>21</td>
<td>83.84</td>
<td>14.83</td>
</tr>
<tr>
<td>No</td>
<td>13</td>
<td>71.2</td>
<td>24.84</td>
</tr>
<tr>
<td>Social validity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>8</td>
<td>69.48</td>
<td>22.46</td>
</tr>
<tr>
<td>No</td>
<td>26</td>
<td>81.93</td>
<td>18.54</td>
</tr>
</tbody>
</table>

N: Number of studies.
X: Mean.
SD: Standard deviation.

PNCD and PEM \( (r = .583, p < .001) \) (Cohen, 1988).

Discussion

The participant and intervention characteristics in the studies in which PRT was used were examined in this study. The effectiveness of PRT in teaching various behaviors to children with ASD was demonstrated. We assessed whether the effectiveness of PRT changed in terms of the characteristics of the intervention. We examined the relationships among the effect sizes calculated by the PND, PNCD, and PEM scores.

When we examined the years in which the studies were published, we observed a linear increase in the number of studies conducted from 1979 to 2012. Thus, the needs of children with ASD in the basic skill areas may have become more prominent in recent years, and PRT awareness may have increased since the 2000s (Bozkus-Genc & Vuran, 2013). Most of the studies on PRT were conducted with children with ASD aged 0–9 years (91.2%). In addition, the majority of the participants were...
Male (72.2%). Children’s early years are the foundation of their future. A growing body of literature suggests that early intensive intervention may greatly enhance outcomes for children with ASD (Vismara & Rogers, 2007). In addition, statistics show that autism is four to five times more common among boys than girls (Centers for Disease Control and Prevention, 2012). Therefore, it is reasonable that the majority of the participants were young (Koegel et al., 2012) male (Bozkus-Genc & Vuran, 2013; Masiello, 2007) children.

PRT was applied in various settings, such as clinics, schools, homes, community settings, or multiple settings, and by professionals, parents/caregivers, or peers. Whereas children with ASD may direct their attention to certain characteristics of stimuli in their environments, they may completely ignore others (Lovaas, Schreibman, Koegel, & Rehm, 1971). Consequently, their generalization skills might be negatively affected. It is suggested that researchers utilize multiple settings and interveners in order to increase the learning experiences of children and enable the generalization of the newly learned skills (Dunlap, Koegel, & Burke, 1981; Lovaas et al., 1979). In studies of PRT, skills appropriate to the needs and ages of children with ASD, such as interaction skills, social skills, play skills, and academic skills, were targeted. ASD is characterized, in varying degrees, by difficulties in social interaction, verbal and nonverbal communication, and repetitive behavior. Therefore, it can be suggested that participants’ core deficiency skills were targeted, which may have positively affected the social validity of the studies.

Inter-observer reliability data were collected in all of the studies. However, treatment integrity data were collected in only 44% of the studies. Accurate measurement of the dependent and independent variables is an important prerequisite to experimentally establish the existence of a functional relationship. This practice is also important for purposes of external validity and the replication of the procedures used within a study (Wheeler, Baggett, Fox, & Blevins, 2006). Therefore, future research studies are needed to demonstrate the treatment integrity of PRT. Maintenance and generalization data were collected in only half of the studies. However, like all students with disabilities, the maintenance and generalization of newly acquired behaviors are crucial for children with ASD. Moreover, maintenance and generalization are key elements to increase the instructional efficacy, allowing for the allocation of more time to teaching new behaviors. Although social validity is also critical, social validity data were collected in only one-quarter of the studies. This ratio is very low for studies in which the core skills of children with ASD were targeted (Bozkus-Genc & Vuran, 2013; Masiello, 2007).

Quality indicators such as maintenance, generalization, and social validity data were not collected in some studies. The lack of these quality indicators decreased the quality of the studies (Horner et al., 2005), and may pose a limitation to the some previous studies.

This study found that PRT was effective in teaching various behaviors to children with ASD. Based on an examination of the mean effect sizes of the studies included in this meta-analysis, calculated by three methods, we conclude that PRT is fairly effective (Ma, 2006, 2009; Scruggs & Mastropieri, 1998, 2001). The current findings support previous research that individually demonstrated the

### TABLE 5

<table>
<thead>
<tr>
<th></th>
<th>PND</th>
<th>PNCD</th>
<th>PEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>PND Pearson’s r</td>
<td>.749</td>
<td>.598**</td>
<td></td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>284</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PNCD Pearson’s r</td>
<td>.749**</td>
<td>.583</td>
<td></td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>284</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PEM Pearson’s r</td>
<td>.598</td>
<td>.583**</td>
<td></td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>284</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

PND: Percentage of nonoverlapping data.
PNCD: Percentage of nonoverlapping corrected data.
PEM: Percentage of data points exceeding median.
**Correlation is significant at the .01 level (2-tailed).
effectiveness of PRT in teaching various behaviors to children with ASD (Jones & Feeley, 2007; Koegel et al., 2012). In practice and future research studies, PRT can be used to teach pivotal behaviors to children with ASD. When the individual effect sizes of the 34 studies were examined in terms of PND scores, 13 studies were highly effective, 11 were fairly effective, and 10 were questionable or ineffective. In terms of PNCD scores, of the 33 studies, 14 were highly effective, nine were fairly effective, and 10 were questionable or ineffective. In terms of PEM scores, of the 34 studies, 27 were highly effective, four were fairly effective, and three were questionable or ineffective. When the calculation methods were compared, it was found that the PND and the PNCD scores yielded similar results, whereas these two methods were highly differentiated from the PEM scores. This result is consistent with the findings of Wolery et al. (2010). Visual analysis of the graphs of the studies in which the effect size calculations differed greatly revealed a leap or fluctuation in the data points at the baseline level; these leaps or fluctuations may have caused this difference in the calculations (Faith et al., 1996; Ma, 2006; Scruggs & Mastropieri, 1998). In the literature, researchers have mentioned that different effect size calculation methods may lead to different conclusions about the degree of treatment effectiveness for the same data set (Parker et al., 2007).

In this study, the PND, PNCD, and PEM scores were examined in terms of different variables; however, the only difference found was in terms of dependent variables. While the PND scores were questionable or ineffective in terms of play skills or social skills, PNCD scores were only questionable or ineffective in terms of play skills. The PEM scores were fairly or highly effective in terms of all of the dependent variables. This finding supports the fact that PRT was especially developed for teaching pivotal areas such as communication and interaction skills (Koegel, Koegel, Harrower, & Carter, 1999). However, it is important to consider the fact that in this meta-analysis, a few studies targeted play skills and social skills. Importantly, the data of these studies were not analyzed by inferential statistics; instead, they were analyzed by using only descriptive statistics. Inferences based on the results of a few studies analyzed by only descriptive statistics may not be accurate.

Correlation coefficients calculated to determine the degree and the direction of the relationship among the PND, PNCD, and PEM scores indicated that there was a significant relationship among these three calculation methods; the relationship between PND and PNCD scores was stronger than the relationship between PND and PEM scores and PNCD and PEM scores (Cohen, 1988). This outcome was expected, since these three methods are based on overlap or the same theoretical foundation (Wolery et al., 2010). PNCD and PEM scores were strongly and positively correlated with PND scores, even though they were developed to overcome the limitations of calculating PND scores reported to be insufficiently sensitive. Thus, because it is easier to calculate PND scores than PEM and PNCD scores, the calculation of PND scores may be preferred for effect size calculations.

PNCD is an effective method to deal with trends. Therefore, it can be used in situations when pre-intervention measurements do not show pure random fluctuations (Malonov & Solanas, 2009). In this meta-analysis, PND and PNCD scores differed because there was a trend in the baseline for a few studies; however, this difference was not statistically significant. Therefore, when there is no trend, PND scores might be calculated instead of PNCD scores, since it is very difficult to digitize the data and calculate PNCD scores.

This study has some strength. First, quantitatively synthesizing SCR studies is a laudable goal because it can increase the objectivity of syntheses and allow the quantification of the potential effects (Wolery et al., 2010). Therefore, the current study may support the existing literature about PRT and teaching children with ASD. Second, PNCD is a method developed to omit pre-intervention trends from the data. Thus, PNCD can be used when the data are sequentially dependent on each other, which allows for the calculation of effect sizes using all of the data points in both baseline and intervention phases. However, PNCD scores were used in a limited number of studies. Therefore, the present study might have contributed to the related literature. Third, this study was conducted to compare
the different effect size calculations for SCR designs. Therefore, the results of the present study might provide clues and guide experts from related fields.

Several limitations need to be addressed within the framework of this meta-analysis. This study was limited to studies in which SCR designs were used because studies in which group experimental designs were used were excluded. In future meta-analyses, studies in which group experimental designs were used to test the effectiveness of PRT might be examined. In the current study, the quality of the studies included was not assessed, and the effect size calculations were not assessed in terms of the quality of the studies. However, it is crucial to examine the relationship between the quality and effect size of the studies. Therefore, new studies might be designed in order to examine the quality of the studies and the relationship between the quality and effect size of the studies. PNCD may not be a preferred method because the data points must be digitized, the calculations are very difficult, there are no criteria that can be used to interpret the scores, and it can give misleading scores when there is a trend in the data (Manolov & Solanas, 2009). Therefore, a software program might be developed to facilitate the calculation of PNCD scores. As the number of studies was limited, descriptive statistics were used in this study to examine the effect sizes in terms of the different variables. However, future analyses should examine the effect sizes of more studies by using inferential statistics in terms of different variables. The overlap methods are not an estimate of the magnitude of the effects between conditions, although they are meant to represent magnitude. The magnitude of the data (i.e., effect size) can be quite different. Overlap methods report only the proportion of overlap across the conditions. Thus, they do not provide an estimate of the magnitude of effects (Wolery et al., 2010).

References


Odom, S. L., Boyd, B. A., Hall, L. J., & Hume, K.
National Professional Development Center on Autism
Masiello, T. L. (2007). Effectiveness of pivotal re-
Manolov, R., & Solanas, A. (2009). Percentage of
Ma, H. H. (2006). An alternative method for quant-
Ma, H. H. (2009). The effectiveness of intervention
Koegel, L. K., Koegel, R. L., & Schreibman, L.
Koegel, L. K., Koegel, R. L., & McNe-
Koegel, L. K., Koegel, R. L., & Schreibman, L.
Koegel, R. L., Koegel, L. K., & McNer-
Koegel, R. L., Koegel, L. K., & Mc-
Koegel, R. L., Schreibman, L., Koegel, R., & Rehm,
Lovaas, O. I., Schreibman, L., Koegel, R., & Rehm,
Lovaas, O. I., Koegel, R. L., & Schreibman, L.
Laski, K. E., Charlop, M. H., & Schreibman, L.


Received: 15 December 2014
Initial Acceptance: 12 February 2015
Final Acceptance: 16 April 2015
An Evaluation of App-Based and Paper-Based Number Lines for Teaching Number Comparison

Pei-Lin Weng
William Paterson University

Emily C. Bouck
Michigan State University

Abstract: Number comparison is a fundamental skill required for academic and functional mathematics (e.g., time, money, purchasing) for students with disabilities. The most commonly used method to teach number comparison is number lines. Although historically paper number lines are used, app-based number lines may offer greater flexibility. This study compared using paper-based and app-based number lines to teach number comparison skills to students with intellectual disability using a single-subject alternating treatments design. Three secondary students with intellectual disability participated in a classroom setting during baseline and intervention phases and a simulated grocery store setting during generalization phases. Both the adapted paper-based and app-based number lines were effective for students with intellectual disability. However, the app-based number line was slightly more effective in terms of accuracy and completion time during number comparison and price comparison tasks.

Arabic number comparison is an important skill for students with disabilities – used in academic mathematics, other subject areas (e.g., science – comparing temperature, and history – comparing event timelines), and daily living skills (i.e., grocery shopping; Holloway & Ansari, 2009; Storey & Miner, 2011). Number comparison involves comparing the magnitudes of number representations, which can be non-symbolic numeral representation (two dots vs. three dots) or symbolic numeral representation (e.g., one vs. three or 1 vs. 3). The ability to compare magnitudes of Arabic numerals is important because Arabic numerals are the most common numerical representation present in one’s daily life (Cohen, Kadosh, & Walsh, 2009).

In order to perform number comparison independently, an individual must master the following skills: recognizing numerals, recognizing the association between numeral representations and magnitudes, and establishing a systematic understanding of language associated with number comparison (e.g., smaller, bigger) and the sequence of numerals (Confer, 2005; Faulkner & Cain, 2009). Students who master number comparison typically demonstrate a concept called mental number lines (Anghileri, 2006; Mosley, 2001). Mental number lines are mental representations of Arabic numerals, which are organized horizontally and sequentially (Ansari, Holloway, Price, & Van Eimeren, 2008; Sousa, 2008).

There are many approaches to teaching number comparison. Among these approaches, number lines – which provide visual cues of numerical ordering relationships – are the most common method used from pre-K to grade 8 (Confer, 2005). Typically developing students acquire a mental number line at approximately 8 years of age (Okamoto, 2010). For students with disabilities, and specifically students with an intellectual disability, number comparison remains a challenge (Sandknop, Schuster, Wolery, & Cross, 1992). The difficulty with number comparison can result from a lack of mental number lines, and students who struggle with mental number lines can benefit from using a concrete number line as a visual cue (Confer, 2005; Holloway & Ansari, 2009). Number lines are a common support used in the secondary education class-
room as students with intellectual disability perform functional or academic mathematics skills (Browder, Spooner, & Trela, 2011).

There are various mediums of presenting number lines when teaching students with and without disabilities. Paper is the most common medium to present number lines (Fletcher, Boon, & Cihak, 2010). However, computers can also effectively present number lines (Fuchs et al., 2006; Lin, Podell, & Tournaki-Reid, 1994). As computers are increasingly replaced with more mobile devices (e.g., iPad), the potential for numbers lines to be placed on such mobile devices through apps also exists (Herbert, 2010). Mobile devices not only offer the capacity to present instructional materials much like traditional computers, but also provide advantages, such as portability, long battery life, multi-touch capacity screens, and ease of access to numerous apps (Clark & Luckin, 2013; Douglas, Wojcik, & Thompson, 2012). The portability and long battery life of these mobile devices allow students to gain access to instructional materials in multiple settings and for multiple tasks (Vedantham & Shanley, 2012).

Although mobile devices and their apps are presumed to provide many benefits to students with disabilities, research on their potential is lacking. The majority of research on mobile devices focuses on the provision of video modeling to students with disabilities with these devices, rather than evaluating the effectiveness of apps for learning and skill acquisition. However, within the existing limited literature, support exists for use of apps on mobile devices to support individuals with disabilities, including individual with intellectual disability. For example, Cihak, McMahon, Smith, Wright, and Gibbons (2015) successfully taught individuals with intellectual disability to use an email app on an iPad to communicate; Cihak et al. also taught the individuals to send an email via desktop and laptop computers, providing the individuals choice and flexibility of technology medium. In another study, Hsu, Tang, and Hwang (2014) successfully taught three secondary students with moderate intellectual disability to use an app on a mobile device to support purchasing skills.

The purpose of this study is to compare the effects of paper-based number lines and app-based number lines to teach number comparison skills for students with intellectual disability. In this study, the authors compared paper-based and app-based number lines – presented on an iPad – to perform number comparison for students with intellectual disability. This study systematically compared the effectiveness of two number lines for teaching number comparison. Research questions in the current study include the following: (a) Can the effectiveness of number lines be replicated in teaching number comparison for students with intellectual disability?, (b) Are app-based number lines as effective as paper-based number lines in teaching number comparison for students with intellectual disability?, and (c) What are students and practitioners’ perspectives on learning and teaching number comparison via a mobile device?

Method

Participants

Three students with intellectual disability were recruited from a high school in the Midwest. All participating students met the following criteria: (a) identified with a moderate intellectual disability, (b) educated in high-school, (c) Individualized Education Program (IEP) goals related to academic and functional mathematics, (d) ability to identify numbers 0–10, (e) visual and hearing functioning within normal limits before or after correction, and (f) lack of mastery of single-digit number comparison. The first five criteria were judged based on teachers’ reports. The last criterion was assessed using a five-question pre-assessment, which was given to determine the lack of mastery of single-digit number comparison if the students selected target numbers with 0–80% accuracy. All students received academic instruction in a regular classroom at rates less than 40% of the day and functional instruction in the special education classroom at rate of more than 60% of the day.

Seth. Seth (pseudonym) was a 15-year-old, 10th-grade, male student with a moderate intellectual disability. Seth had a Full Scale IQ of 51, Nonverbal IQ of 56, and Verbal IQ of 51, according to the Stanford-Binet Intelligence Scales Fifth Edition (SB-5; Falvo, 2005). His
general adaptive composite score of 69 according to Adaptive Behavior Assessment System (ABAS-II) indicated a moderate disability (Lichtenberger & Kaufman, 2009). Seth was sociable and outgoing, but he also had a short attention span and became easily distracted by others and environmental stimuli. His teacher reported his performance fluctuated daily, and that Sean loved to use iPads.

**Helen.** Helen (pseudonym) was a 14-year-old, 9th-grade, female student with a moderate intellectual disability. Helen had a Full Scale IQ of 52, Nonverbal IQ of 57 and Verbal of IQ 52 according to SB-5 (Falvo, 2005). Her general adaptive composite score of 40 according to ABAS-II indicated a moderate disability. Helen was typically friendly with staff and peers and she liked to use iPads in school.

**Kelsey.** Kelsey (pseudonym) was a 14-year-old, 10th-grade, female student with a moderate intellectual disability. Kelsey had a Full Scale IQ of 42, Nonverbal IQ of 46, and Verbal of IQ 43, suggesting a moderate intellectual disability according to SB-5 (Falvo, 2005). Her general adaptive composite score of 41 according to ABAS-II was consistent with a moderate disability. Kelsey was friendly, outgoing, and eager to please teachers; she also liked to be around her peers. The teacher reported Kelsey had experience using iPads in class.

**Settings**

The study was conducted at a public junior and senior high school in a small room adjacent to students’ main classroom. The small room contained a desk and three chairs. On one wall were a washer and dryer, and on the other wall a storage cabinet. Using the adjacent room for sessions minimized possible distractions from activities in the main classroom. The room arrangement was consistent in the baseline and intervention phases. For the generalization phase, the researchers set up a simulated grocery store in the room. Five pairs of grocery products, each with two different price tags, were set on the table.

**Materials**

**iPad.** An Apple iPad 2 was used in the study. The size of an iPad 2 – $9.5 \times 7.31 \times 0.34$ inches – provided sufficient space to display the number line app used in the study. The participating students were familiar with iPads because their classroom had two iPads.

**Number line app.** The number line app used in the number comparison study was developed by Authors (2015), who employed theoretical and empirical guidance for developing software and instructional technology for students with disabilities (see Higgins, Boone, & Williams, 2000; Tammaro & Jerome, 2012; Walker, 2011; see Figure 1). The number line app was age appropriate for secondary students and had a consistent layout, simple design, non-disturbing items, and easy navigation. At the bottom of the app was a horizontal number line consisting of whole Arabic numerals from 0 to 10 printed in the same-sized font. In the top, left corner were three control buttons: refresh, marker, and eraser icons. When the marker function was on, students used their fingers to circle numerals. When numerals were circled (e.g., 2 and 6), arrays of dots corresponding to the circled numerals appeared above the numerals (e.g., 2 dots and 6 dots). With the eraser function was on, the students used their fingers to erase the circles.

**Paper-based number line.** The paper-based number line was presented on a laminated letter-size paper. The lamination allowed students to use markers. The number line – located on the bottom of the paper – consisted of a horizontal line with whole Arabic numerals from 0 to 10 printed in the same size. Additionally, there were 11 same-sized strips ($6.5” \times 0.8”$) with different amounts of red dots ranging from no dots to 10 red dots (see Figure 1). The strips were held by the researcher, who added individual strips to the number line when the student circled numerals using a marker. The strips were removed from the number lines by the researcher as soon as the students erased the circles on the number lines using a dry erase board eraser.

**Assessment.** There were 15 number comparison assessments throughout the baseline and intervention phases and 6 price comparison assessments during the generalization phase. Each number comparison assessment consisted of five questions. The assessments were randomly generated using the website Random.org. Each question contained two whole Arabic numerals between 0 and 10. To
minimize the order effect, the arrangement of the two numerals with different quantities in each question set was randomized. A coin flip was used to determine the order of the two numerals (e.g., the bigger number on the right and the smaller number on the left [e.g., 5 vs. 8]; the smaller number on the right and the bigger number on the left [e.g., 5 vs. 3]), while ensuring that no more than two consecutive sets had the same order of numbers.

Figure 1. Two types of number lines.
Thus, for each assessment – one used per session – only two or three question sets had larger numbers to the right of smaller numerals. Each of the questions were printed on a piece of paper (2” x 8”) in Times New Roman, size 48 font, and presented to each student one-by-one.

**Grocery products.** During generalization, a total of 25 pairs (i.e., 50 items) of grocery items were purchased based on the teacher’s suggestions of items commonly purchased by the students during their weekly shopping. Each pair of the items – containing two brands of the same type of product (e.g., Dawn Dish Soap and Method Dish Soap) – was assigned two different single-digit prices (e.g., $3 and $5). The first author randomly selected five of the 25-paired items for each of the six generalization sessions, each consisting of five trials. Some of the items were repeated twice, but none of the items were repeated more than twice.

**Dependent and Independent Variables**

The independent variable was the use of two types of number lines (i.e., app- and paper-based number lines) in conjunction with a system of least-to-most prompts and a constant time delay for all tasks analyses except for the steps, “say or point to the target number” and “say or point to the lower price tag”. The dependent variables were (a) the percentage of correctly answered number comparison questions per session and (b) task completion time per session.

**Design and Procedures**

A single-subject alternating treatments design study was employed to compare the effects of the two types of numbers lines on teaching number comparison to three students with moderate intellectual disability. An alternating treatments design was selected because it allows comparing two intervention conditions by rapidly alternating instructions (Barlow, Nock, & Hersen, 2009). This alternating treatments design study consisted of baseline, intervention, and generalization phases.

**Baseline.** During each of the five baseline sessions, students worked with the first author one-on-one. Students were presented with five questions, each on a separate piece of paper, one at a time. After giving the student a question, the researcher verbally asked one of the directions: (a) “Which one is bigger?” or (b) “Which one is smaller?” The selection of either oral direction was randomly selected and no more than two consecutive trials with the same direction and no same direction was asked more than three times. All students went through the same order of the oral directions. Two behaviors were recorded: (a) time to complete five questions per session, and (b) percentage of correctly answered questions (either bigger or smaller).

**Training.** During training, the researcher individually taught students how to use the two number lines using a six-step task analysis (see Table 1). Students were first given the paper-based number line, a marker, and an eraser. Students were presented with a pair of numbers on a paper sheet and asked to circle

### Table 1

<table>
<thead>
<tr>
<th>Task Analysis for Intervention</th>
<th>Task Analysis for Generalization</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Circle 1st number on a number line</td>
<td>1. Circle 1st number on a number line</td>
</tr>
<tr>
<td>2. Circle 2nd number on a number line</td>
<td>2. Circle 2nd number on a number line</td>
</tr>
<tr>
<td>3. Say/point to the target number*</td>
<td>3. Say/point to the target number*</td>
</tr>
<tr>
<td>4. Circle on the paper problem</td>
<td>4. Say/point which price is smaller*</td>
</tr>
<tr>
<td>5. Erase 1st circles on a number line</td>
<td>5. Say/point which price is cheaper*</td>
</tr>
<tr>
<td>6. Erase 2nd circles on a number line</td>
<td>6. Erase 1st circles on a number line</td>
</tr>
<tr>
<td>7. Erase 2nd circles on a number line</td>
<td>7. Erase 2nd circles on a number line</td>
</tr>
</tbody>
</table>

* Do not provide least-to-most prompting.
the two target numbers with the marker on the number line. Within two seconds of when students circled one number on the number line, the researcher placed a strip with an array of dots corresponding to the numbers above the number line. After circling two numbers, students were asked which was smaller (or larger). After students answered and circled the target number on the paper, the researcher instructed students to erase the circles using the erasers and the researcher removed the two strips one by one. The students were given another two opportunities to practice these steps. After two practices, students were given another question sheet to determine the training outcomes. If students reached the criterion of 83.3% accuracy on one question (i.e., five out of six steps, which include manipulating the number line and answering questions; see Table 1), the training for using paper-based number line was considered concluded. All three students reached 100% of accuracy of paper-based number line after the first trial.

Next, students were shown the app-based number line, using the same procedure as with the paper-based number line. Students were first introduced to the app’s marker and eraser icons. Students were then given a piece of paper with a pair of numbers and asked to circle the target numbers. Seth and Helen operated the number line app using their fingers, while Kelsey operated the app with a stylus because the iPad screen was not 100% responsive to her fingers. Once students circled the numbers, the arrays of dots appeared immediately and automatically. After students selected the target number from the number line and then the question sheet, the researcher asked them to press the erase icon and use their fingers or stylus to erase the circle. Once the circles were erased, the dots disappeared immediately. Finally, students were given two more opportunities to use the app-based number line. The same mastery criterion as with the paper-based number line was applied. Seth and Helen reached 100% of accuracy after first trial and Kelsey reached 83.3% accuracy after the second trial using a stylus.

Intervention. The intervention phase consisted of 10 sessions – five for each condition (paper-based and app-based number line); students worked one-on-one with the first author. The sequence of the two conditions alternated randomly, with no more than two consecutive sessions of the same condition. During each session, students were given five number comparison questions. After the two numbers were circled on either number line – and their resulting visual array appeared, the researcher verbally asked the student which number was either smaller or larger, as in baseline. Each session concluded once students had completed five questions. The researcher recorded the same target behaviors as in the baseline condition.

Generalization. The generalization phase consisted of six sessions – three for each intervention condition. The sequence of the two conditions randomly alternated, so no more than two consecutive sessions were the same condition. During the generalization phase, participating students used either type of number line to select the lower-priced grocery item in a simulated grocery store in the classroom. The procedures in the generalization phase were similar to those in the intervention phase, except the tasks were comparing pairs of different, single-digit prices ($2 vs. $4). Each session consisted of five trials of selecting lower-priced items.

Social Validity

Social validity interviews were conducted to assess the use of number lines in learning number comparison and to compare the two types of number lines. Participating students and their special education teacher were interviewed for approximately 10 minutes after the generalization phases. Students and the teacher were each asked nine questions, including open-ended and yes/no questions. The questions for the students included: (a) What do you think about learning to compare numbers?; (b) What do you think about learning to compare prices?; (c) Do you think learning how to compare on a number line will help you solve some questions?; (d) Considering both math questions and comparing prices, what do you think about using the number line?; (e) What do you think about using the number line on an iPad?; and (f) Which number line did you like using when you go shopping and need to compare
prices? The nine social validity for teachers were: (a) What are the advantages of learning number comparison?; (b) Would you teach number comparison in the manner used in this study?; (c) Do you think your students benefitted from learning number comparison skills?; (d) Do you think your students benefitted from using a number line to learn price comparison skills?; (e) Will you continue to use a number line to teach number-related skills?; (f) What did you like about using a number line presented on paper?; (g) What did you not like about using a number line presented on paper?; (h) What did you like about using a number line presented on an iPad? and (i) What did you not like about using a number line presented on an iPad? Interviewees’ were encouraged to elaborate on each question.

Interobserver Agreement and Treatment Integrity

Interobserver agreement was evaluated for 33.3% of sessions for each student. The first author served as the main data collector and three other people served as the second observers, including two special education researchers and one paraprofessional. The second observers alternated observing a total of seven sessions for each participant, including two sessions during the baseline phase, three sessions during the intervention phase, and two sessions during the generalization phase. These sessions in each phase were selected randomly. Interobserver agreement on the percentage correct was 100% for Seth and Helen, and 99.8% for Kelsey. Treatment integrity was assessed by the first author and the second observers for 42.9% sessions (i.e., three sessions each in baseline, treatment, and generalization phases) using a checklist. The items in the checklist included the following: (a) researcher places a number line (and marker and eraser if paper) on the desk; (b) researcher starts to time a session; (c) researcher gives a paper problem and provides time delay, (d) if paper, researcher puts down matching array of dots and provides time delay after the student circles the first number on the number line; (e) if paper, researcher puts down matching array of dots and provides time delay after the student circles the first number on the number line; (f) researcher points to arrays and asks which one is bigger/smaller followed by time delay; (g) researcher takes away paper problem followed by time delay; (h) researcher takes away dots if using the paper number line; and (i) researcher stop timing session. Treatment integrity was achieved 100% for all three students.

Data Analysis

Visual and quantitative data for the two types of number lines in the baseline, intervention, and generalization phases were analyzed. In each session, trial recording technique was used to document the responses for each trial (Ayres & Gast, 2010). For each session, the accuracy of the answers was calculated as a percentage, dividing by five the total of correct selections of either the target numbers or the lowest price; these data were then graphed onto line graphs. The duration recording technique was used to note the completion time for each session to compare the efficiency of the two number lines (Ayres & Gast). Visual analysis techniques were used to analyze level of the graphed data points within and between phases (Gast & Spriggs, 2010; Lane & Gast, 2013). Level stability was determined using the 80%–20% criteria: 80% of data points fall on or within the stability envelope, which is 20% of a median value from a baseline (Gast & Spriggs). The mean accuracy and two types of effect sizes were reported across phases for quantitative data analysis (Maggin, Briesch, & Chafouleas, 2013). First, each student’s mean and range of accuracy in each phase was calculated as a percentage. Second, two effect size indices – standard mean differences (SMD) and regression coefficients – were reported to present the effectiveness of two types of number lines. SMD were obtained using the mean differences between baseline and intervention divided by the standard deviation. Standardized coefficients – one type of regression-based effect sizes – were calculated to present trends and levels separately (Beretvas & Chung, 2008).

Results

For the three students, the visual analysis and SMD indicated both types of number lines were effective. However, the app-based num-
ber line was slightly more effective and resulted in more stable outcomes than the paper-based number line for two of the students. Figure 2 displays the accuracy of the number comparison performed and Table 2 summarizes the mean, level stability, SMD, and standardized regression coefficients for trends and levels, and completion time for each student.

**Seth**

During the baseline phase, Seth correctly answered an average of 76% of the questions, with sessions ranging from 40 to 100%. Seth’s data points showed a variable level as only 60% data points within the stability envelope. During the intervention phase, Seth correctly answered 100% of questions with zero-celerating trend and stable level (100% data points within the stability envelope) under both the paper- and the app-based number line conditions. Both Seth’s SMD for the paper- and the app-based number line conditions were 1.24, indicating both number lines demonstrated positive effects. Regarding completion time during the intervention phase, Seth took 51 seconds longer, on average, to complete all questions per session under the paper-based number line condition than the app-based number line condition (5:44 vs. 4:53). During the generalization phase, Seth selected the lower-priced item with 100% accuracy with stable trend and level (100% data points within the stability envelope) under both number line conditions. He took, on average, 55 seconds longer to complete questions using the app than the paper-based number line condition during generalization.

**Helen**

During baseline, Helen correctly answered an average of 52% of questions, with range from
20% to 60% with stable level (80% data points within the stability envelope). During the intervention phase, Helen performed slightly more accurately using the app-based number line. She correctly answered 96% of questions on average per session under the paper-based number line condition with stable level (80% data points within the stability envelope) but 100% under the app-based number line condition with stable level (100% data points within the stability envelope). Both conditions were effective for Helen judged from effect size indices: SMD (i.e., 1.64 under the paper-based condition; 1.72 for the app-based condition) and the standardized regression coefficient for level (i.e., 1.26 under the paper-based condition, $p < 0.01$; 1.36 for the app-based condition, $p < 0.01$). On average, Helen took 1 minute and 45 seconds longer to complete a session under the paper-based than the app-based number line condition (4:58 vs. 3:13, respectively). During the generalization phase, she selected the lower-priced item with 100% accuracy with a zero-celerating trend and stable level under both number line conditions. She took an average of 1 minute and 12 seconds longer to complete a session under the paper-based than the app-based number line condition.

**Kelsey**

During baseline, Kelsey correctly answered an average of 52% questions, with a range from 40% to 80% with cyclically variable trend (Gast & Spriggs, 2010) and variable level (60%...
data points within the stability envelope). During the intervention phase, she performed better under the app-based number line condition. She correctly answered an average of 80% of questions per session under the paper-based number line condition and 92% under the app-based number line condition. Kelsey’s data points under the paper-based number line condition showed cyclically variable trend and variable level (20% level stability) across sessions. Her data under the app-based number line showed zero-celerating trend and stable level across four sessions (sessions 7, 10, 11, and 14) and then decelerating for the last session (session 15), resulting in 80% level stability. The effect size indices also indicated the superior effect of the app-based number line over the paper-based number line: SMD (1.48 [app] vs. 1.21 [paper]) and the standardized regression coefficient for level (1.09, \( p = .051 \) [app] vs. .46, non significant [paper]).

Social Validity

All the students and the teacher indicated they liked using the number lines to learn and teach number and price comparison. Helen stated the number lines helped her compare numbers. In terms of medium preference, Seth and Helen chose the app-based, and Kelsey chose the paper-based number line. As for likes and dislikes about the paper number line, Seth said he liked that he could draw circles on it. However, he disliked several aspects of it, including that he thought it was hard to circle the numbers, the eraser kept falling to the floor, and he did not like that his hands got dirty while using a marker. Helen expressed similar dislikes. Kelsey did not elaborate on her likes and dislikes about the number lines. The teacher thought the paper-based number line could be used as a teaching tool instead of using iPads; paper is more beneficial for students who experience challenges with fine motor skills. She stated, however, that the paper-based number line could carry a greater stigma for students using them.

Regarding the app-based number line, Seth said he liked everything about it, especially that he could operate the number line using just his finger. Helen liked that she could carry an iPad wherever she went. She thought using the app-based number line helped her solve problems faster, and she liked that did not need an adult sitting next to her as she did under the paper-based number line condition. The teacher believed using iPads offered less stigmatization. Seth, Helen, and the teacher all liked the multifunctional quality of iPads.

Discussion

The main purpose of this study was to compare the effectiveness of two types of mediums used to present a number line to secondary students with intellectual disability: an app-based number line and a paper-based number line. The app-based number line was equally effective as, if not slightly more effective than, the paper-based number line in terms of accuracy and task completion time during number comparison (intervention) and price comparison (generalization) tasks.

The study shows use of a number line was effective in teaching number comparison. Comparing the means of the baseline and treatment phases, the three students were more accurate when using a number line than not using one. The use of a number line benefited even Seth, who achieved 76% accuracy in baseline. By visually examining Seth’s data points, we found his performance on number comparison fluctuated during the baseline phase, with a range from 40% to 100%. However, his performance became stable when he used either number line. Seth consistently achieved 100% accuracy independently under both number line conditions. While we could not isolate the factors causing his performance to fluctuate during the baseline phase, we conclude use of a number line prevented the fluctuation during intervention. Seth successfully executed number comparison tasks.
with the support of a number line; success with number comparison, supported or un-supported, allows students to engage in the application of academics and functional mathematics (Holloway & Ansari, 2009; Storey & Miner, 2011).

The results were inconsistent regarding the average task completion time under duration the app and paper number line conditions. For Seth, his average tasks completion time using the app number line was about 50 seconds faster than using the paper-based number line during the intervention phase, and 50 seconds slower under the generalization phase. For Kelsey, the situation was reversed. Kelsey was about 2 minutes and 45 seconds slower when using the app number line during the intervention phase. However, during the generalization phase, she was about 30 seconds faster, on average, using the app-based number line. Helen completed the tasks faster with the app number line in both intervention and generalization phases. However, it should be noted that the difference in time, especially with regards to the paper number line, could be based on design. For example, with the paper number line the researcher had to place on the dots, and the students had to maneuver multiple objects (e.g., eraser and pen). With the exception of Kelsey during intervention, the differences between the two conditions were all less than two minutes.

Although both interventions were equally effective for Seth, the app-based number line was slightly more effective than the paper-based number line for Helen and Kelsey. The difference between app and paper was relatively minor for Helen during intervention (i.e., 96 vs. 100% accuracy), but larger for Kelsey during the treatment phase (i.e., 80 vs. 92% accuracy) and the generalization phase (i.e., 46.7 vs. 86.7% accuracy). In this study, the app-based number line was presented on a mobile tablet, specifically the iPad. One hypothesis for the difference in performance for the two female students between the app-based number line and the paper-based number line involved the design of the study. The app-based number line simplified the objects to manage during the task, as all were virtual and contained within one device. Under the paper-based number line condition, participants were required to hold three items, including the number line, an eraser, and markers, and to make certain these items did not fall from the clipboard. This was a limitation of the study. The authors felt this hypothesis might be more applicable to Kelsey given her hand motor control was less developed and she tended to drop objects more than the other students, which may have compromised her working memory and attention to the tasks (Low, Jin, & Sweller, 2011), although it did not impact her task completion time during intervention (i.e., faster with paper than app number line during intervention). However, the authors could have, and teachers can, make the paper-based number line more efficient by using dry erase markers with an eraser pad on and end and attaching the marker to a clipboard via string.

Clearly, both number line options provided advantages to students and teachers should consider using at least one, if not both. Both types of numbers line provide advantages and disadvantages to students and teachers, and teachers to consider their context and individual student preferences as well as strengths and challenges when making such decisions. In terms of the app-based number line, it offers greater independence and less adult-required assistance, when using a number line with dots for additional visual cues (i.e., the researcher had to manually apply the dots). Based on the structure of the paper-based number line in this study, the app-based number line provide more convenience with less things to hold as well as the potential for greater portability within community-based settings, including the potential or less stigmatization when using an iPad in a grocery store than a number line on a clipboard. However, mobile devices, such as iPads, are more expensive than make a laminated paper-based number line. In addition, teachers need to be cognizant regarding the fragileness of these devices, particularly as compared to a paper-based number line. While a protective case can help to reduce breakage, a potential concern remains.

As noted, student preference is important, especially when considering technology. Technology abandonment is a real concern for students with disabilities, and taking student consideration into account can reduce
the potential for abandonment, which does not only cost money but also leaves students without assistance (Parette & Scherer, 2004). Two of the students preferred the app-based number line to the traditional medium in this study and one preferred the paper-based number line. Despite the appeal of apps and mobile devices (i.e., more sophisticated technology), student preference in actual learning situations to new technology is not always found (Bouck & Weng, 2014; Woody, Daniel, & Banker, 2010). Educators need to ensure that students are comfortable with the medium of their learning, regardless of low-tech or high-tech.

**Implications for Practice**

There are several practical implications that arise from this study. First, secondary special education teachers should continue to incorporate a number line of any type into their teaching to help students with number comparison. The students were more successful in correctly comparing single digit numbers with either number line condition than without. For students without a consistent mental number line, paper or app-based number lines can support this skill in both basic and applied (i.e., grocery shopping) situations. The second implication is that teachers should not discard a paper-based number line just because the app-based number line presented on the mobile device demonstrated a slightly better effect for two of the students. The paper-based number line was an effective intervention. As the special education teacher indicated in her social validity interview, the paper-based number line should be introduced to students because it is a cheaper and more readily available medium to use in classroom settings. However, teachers may also wish to incorporate innovative ways to presenting a number line, such as using an app-based option. Use of the number line app on a mobile device can offer a more efficient and socially-acceptance means of supporting students in non-classroom settings.

**Limitations and Future Directions**

This research study, like all research studies, involves limitations. The first limitation is the possible multiple-treatment interference inherent in an alternating treatments design. Given the anticipation of the possible treatment interference, researchers applied methodical strategies (i.e., randomizing the sequence of administering the two conditions and administering only one treatment condition per session) to minimize the possible effects. Future studies should add the best treatment phase (i.e., a phase where the superior condition within the intervention phases was applied) to the design. If the superior condition continued to demonstrate its effect during the best treatment phase, one could rule out the interference effect of two treatment conditions. Researchers may also want to conduct an alternating treatments design with a multiple baseline in the future to provide greater confidence in experimental control.

A second limitation is the possibility of novelty effects produced by an engaging tool such as an iPad. The authors considered this possibility (i.e., novelty effect when using iPads) in the design of the study and attempted to limit such an effect by making the two number lines as similar as possible. However, the interactivity of the iPad as well as its socially-desirable nature are difficult to control for when comparing to paper-based alternatives. Another limitation involved that the stimuli used in this study were single-digit numeral comparison using an adapted number line containing only 0 to 10 digits. Learning how to compare single digits apparently will not be sufficient because secondary students with disabilities are more likely to encounter multi-digit numeral comparison in academic settings and the real world. Therefore, future studies need to look into whether the adapted numbers are suitable for comparing multi-digit comparison using numeral comparison strategies, such as the decomposition model.

Finally, both the app-based and paper-based number lines used in this study contained non-symbolic number cues (i.e., arrays of dots). It appears that the non-symbolic number cues were effective in facilitating number comparison tasks; however, the study did not directly investigate the effects of non-symbolic numeral cues by comparing it to a controlled condition (e.g., number line without the non-symbolic number cues). Researchers need to empirically examine the effect of the non-
symbolic number cues by directly comparing number lines with and without these non-symbolic cues for students with intellectual disability. In addition, the non-symbolic number cues only presented to students when a numeral was selected. Future research can also examine the presentation of additional visual cues by comparing: (a) all non-symbolic number cues shown above a number line consistently, and (b) non-symbolic number cues presented only when associated numerals are selected.

References


Authors (2013). iOS number line app.


Received: 3 December 2014
Initial Acceptance: 5 February 2015
Final Acceptance: 1 May 2015
Abstract: Effects of an embedded simultaneous prompting procedure to teach STEM (science, technology, engineering, math) content to three secondary students with moderate intellectual disabilities in an inclusive general education classroom were evaluated in the current study. Students learned discrete (i.e., geometric figures, science vocabulary, or use of technology to publish) and chained tasks (i.e., linear equation, Punnett square, or Internet research) from a peer tutor and paraprofessionals. Using a multiple probe across participants design, results showed students reached criterion in two to eight sessions and maintained the skills for one month following intervention. Generalization was also at higher levels than in the baseline condition. In addition, general education students who attended class with the participants expressed positive comments based on the intervention.

The National Science Board has argued that every student in the United States “. . . deserves the opportunity to achieve his or her full potential” (NSF, 2010, p. v). As STEM (science, technology, engineering, math) permeates every aspect of our lives, it is crucial that all students have access to this content. Students benefit from science content by learning about themselves and the natural world; from engaging in technology by learning how to use smart phones, mobile devices, and computers; from engineering concepts by learning how to solve problems; and from math content by learning how to budget their money and determining how much they have for groceries (Science Pioneers, 2013). There is a national emphasis on preparing students for STEM careers; however, all citizens, even those not pursuing STEM careers, should be able to participate in the scientific and technical issues affecting our society (Matthews, 2007). Traditionally, students with disabilities have been underrepresented in STEM education, and students with moderate intellectual disabilities are especially at risk.

Common Core State Standards (CCSS) and the Next Generation Science Standards (NGSS) emphasize the importance of high expectations for all students. According to Kleinert, Kearns, and Kennedy (1997), “. . . one of the keys to ensuring high expectations for every child is requiring that all students be included in measures of educational accountability” (p. 88). Based on these assertions, it is likely that assessments for all students, including alternate assessments based on alternate achievement standards (AA-AAS) for students with moderate and severe disabilities (MSD), will be derived from the content recommended in the CCSS and NGSS.

While practices for implementing the AA-AAS vary from state to state, having access to the general education curriculum is key for promoting student progress in content areas. Jackson, Ryndak, and Wehmeyer (2010) argued that the interplay between context of...
instruction and curriculum content is a critical consideration in student learning and progress since students with disabilities who are included in general education contexts are more likely to engage in activities related to grade level standards than students who are in self-contained settings.

Hudson, Browder, and Wood (2013) conducted a literature review to evaluate evidence-based practices for students with MSD in inclusive contexts. Authors found embedded instruction (specifically using constant time delay [CTD]) to be an evidence-based practice. Embedded instruction can be defined as an instructor (teacher, paraprofessional, or peer) distributing trials throughout a lesson or naturally occurring routine in the inclusive classroom. Studies in the review focused on teaching reading and writing skills (e.g., Collins, Hall, Branson, & Holder, 1999), social studies skills (e.g., Collins, Evans, Creech-Galloway, Karl, & Miller, 2007), science skills (e.g., Jimenez, Browder, Spooner, & DiBiase, 2012), and math skills (McDonnell et al., 2006) to students with moderate and severe disabilities within general education classrooms. Hudson et al. (2013) commended researchers for evaluating procedures to teach a range of academic content in general education settings; however, they recommended additional research on strategies to teach complex and chained skills in inclusive environments. Further, only two quality studies in the review used simultaneous prompting (SP), suggesting a need for further research on this strategy in general education settings.

In most of the quality studies reviewed in the Hudson et al. (2013) review, researchers used systematic instruction to teach students core content in general education settings. Methods of systematic instruction include the system of least prompts (SLP), time delay, and simultaneous prompting procedures. Perhaps the easiest and least time consuming method to implement is the SP procedure because the instructor presents the stimulus and the controlling prompt at the same time on each training trial, conducting daily probe trials prior to instructional trials to determine if the student has acquired the target skill (Collins, 2012). Collins et al. (2007) used the SP procedure to teach core content vocabulary to 4 students with MSD at the elementary, middle, and secondary levels, finding the procedure to be effective whether it was delivered in a segregated or an inclusive classroom setting. Riessen, McDonnell, Johnson, Polychronis, and Jameson (2003) and Head, Collins, Schuster, and Ault (2011) compared the efficiency of the CTD and SP procedures in teaching academic content (e.g., science and history vocabulary, state capitols) to middle or high school students with disabilities. In both studies, the researchers found the CTD and SP procedures to be equally effective, but the SP procedure was found to be slightly more efficient (i.e., faster acquisition of skills).

A body of literature has emerged suggesting that, in addition to academic gains, inclusive settings offer the opportunity for positive social effects on students, both with and without disabilities (e.g., Kennedy, Shukla, & Fryxell, 1997). Results have indicated greater social benefits for students who were included full time in general education classrooms. For example, students who were included full time engaged in more frequent interactions with peers without disabilities, more social contacts across a wide range of activities and settings, higher levels of social support behaviors, larger friendship networks, and longer lasting relationships with peers.

While prior research has shown promising effects as a result of teaching in inclusive classrooms, there are several limitations to the research. For example, few studies to date have taught students complex or chained skills in inclusive settings, and no studies exist evaluating STEM-related chained tasks. Hudson et al. (2013) recommended future research using SP in inclusive settings, and despite previous calls for research, sight word identification remains the skill most frequently addressed in these settings. In addition, no research exists on teaching STEM skills identified as part of the AA-AAS in a general education classroom using an embedded SP procedure. The goal of this study is to evaluate the effects of using an embedded SP procedure to teach STEM-related tasks to students with moderate intellectual disabilities in an inclusive setting.

The current study examined instruction on the state standards related to STEM content on which students taking part in the AA-AAS were tested. The study occurred in a suburban high school in a southern state, therefore fo-
cusing on the state’s AA-AAS. The special education teacher and the general education teacher collaborated to develop a plan for instruction in the general education classroom to be delivered by a paraprofessional or a peer tutor, which included embedding the SP procedure into the natural routine of the general education classroom. Specifically, the current study focused on the following research questions: (a) Will the embedded SP procedure result in the acquisition of STEM content skills by students with moderate intellectual disabilities in the general education classroom? (b) Will peer tutors and paraprofessionals implement the embedded SP procedure with high levels of procedural fidelity? (c) Will the target students maintain the acquired skills over time? (d) Will the target students generalize the STEM skills to situations that simulate the AA-AAS? and (e) Will inclusion of the target students in the general education classroom result in positive social relationships with their peers without disabilities?

Method

Participants

Students. Three students with moderate intellectual disabilities from a public suburban high school participated in the investigation. The students each attended at least two general education classes (some with support from paraprofessionals or peers) and attended the rest of their classes in a self-contained special education classroom. Reading and math achievement scores are not available, as the teacher who implemented the study is no longer at the school.

Kate was a 17-year-old female with multiple disabilities (i.e., visual impairment, moderate intellectual disability) who received services under the educational classification of functional mental disability (FMD; the state’s term for MSD) and participated in the AA-AAS. As a 10th grader, she was to be tested in math and writing for that school year. According to the Wechsler Intelligence Scale for Children: 4th Edition (WISC-IV; Wechsler, 2003), Kate’s full scale IQ was 56. Kate could independently add and subtract two digit numbers and could multiply and divide using a calculator. She worked on functional math skills, including counting money and making purchases. Her IEP goals included engaging in appropriate communication with peers and adults, making purchases in the community, identifying the meaning of weekly vocabulary words, answering comprehension questions, writing with correct grammar, using correct punctuation and capitalization, and completing vocational tasks in school and the community. Kate’s long term goal is to work in a restaurant. Prior to the start of the study, Kate attended a general education choir class and Art class.

Ben was a 16-year, 7-month old male with a classification of multiple disabilities (i.e., autism, moderate intellectual disability). He participated in the state’s AA-AAS and, as a 10th grader, was to be tested in math and writing for that school year. According to the WISC-IV, Ben’s IQ was 47. He could write basic sentences and had strong communication skills, although he was often off subject or talked to himself. Ben’s IEP goals included using appropriate language and topic when communicating with peers and adults, making purchases, identifying the meanings of vocabulary words, answering comprehension questions, writing with appropriate grammar mechanics, completing a vocational task, and using appropriate communication skills. Ben has volunteered in a library and his career goal is to work in one upon graduation. Before the study began, Ben was enrolled in a general education computer applications class (the same class used for the study).

Jacob was a 17-year, 8-month old male identified as eligible for services under the category of FMD. He participated in the state’s AA-AAS and, as an 11th grader, was to be tested in science for that school year. According to the Stanford-Binet Intelligence Scale: Fourth Edition (Thorndike, Hagen, & Sattler, 1986), Jacob’s IQ was 53. He had strong basic reading and comprehension skills and wrote legibly, although he had difficulty writing complete sentences. His IEP goals included using age-appropriate language to communicate, budgeting money, identifying the meanings of vocabulary words, answering comprehension questions, writing using correct grammar, and completing vocational tasks. Jacob’s post secondary goal is to
work with animals in some capacity (e.g., animal shelter). Prior to the study, Jacob attended a general education Drama class.

Others. The special education teacher (first author) collected all baseline, probe, and procedural fidelity data. She was a graduate student in special education, with an undergraduate degree in special education - MSD. She had 5 years of experience teaching in self-contained classrooms and had conducted professional development trainings on systematic instruction for the school district.

The special education teacher selected two paraprofessionals to implement the SP procedure within the general education classroom. The first paraprofessional was a 57-year-old female with a Bachelor’s degree in retail management and marketing. She had 5 years of experience in an FMD classroom, had received extensive training in systematic instruction procedures, and regularly collected classroom data. She had 4 years of experience supporting students in an inclusive related arts class, but did not have any experience using systematic instruction in an inclusive setting. In the current study, she attended the Biology class with Jacob. The second paraprofessional was a 45-year old female who was working toward a degree in human services. She had 1 year of experience in an FMD classroom and in supporting students in inclusive settings (i.e., core content and related arts classes). She had minimal training in systematic instruction and data collection procedures, but did not have any experience implementing systematic instructional procedures in these settings. In this study, she implemented the SP procedure with Ben in the Computer Applications class. The special education teacher also selected one peer tutor to serve as an instructor in the general education setting. The peer tutor was 17 years old, was in the 11th grade, and had a 3.3 grade point average. The peer tutor had 1 1/2 years of experience in the special education classroom and had prior experience using a CTD procedure. She had no previous experience supporting students in an inclusive setting, but in the current study, she attended the Algebra class with Kate.

Prerequisite Skills

The students met the following prerequisites: (a) visual acuity to see stimuli, (b) auditory skills to listen to verbal directions, (c) verbal skills to repeat controlling prompt, (d) ability to engage in task for 10 min, (e) participation in the AA-AAS during the current school year, and (f) daily participation in general education classes. The study received the appropriate Institutional Review Board (IRB) approval from all appropriate agencies and participants.

Target Skills

The special education teacher selected the target skills for this investigation based on students’ IEP goals, relationship to post secondary goals, consideration of AA-AAS standards, and were developed in collaboration with the general education teachers. For example, Kate’s target skills of using linear equations and identifying geometric shapes relate to her goal of working in a restaurant (e.g., linear equations may help her determine how much money she will make at the end of the week, her co-workers may use the names of geometric shapes when referring to items used in a restaurant, geometry is beneficial when considering how to fold napkins, cut pies, etc.). Since Ben is interested in working at a library in the future, looking up information on the internet, as well as learning vocabulary words about ways to publish on the internet, would both be integral skills to making his transition from school to post school employment successful. Jacob wants to work with animals in some capacity in the future and was a member of the recycling club at school. These factors were taken into consideration when selecting his goals related to the vocabulary words on homeostasis and Punnett squares. An expert in STEM was consulted to ensure the tasks were aligned to STEM content. Information on the specific targeted discrete and chained skills for each student is listed in Table 1.

Setting

The special education teacher conducted baseline probe, daily probe, maintenance,
and generalization probe sessions in the special education classroom. Instructional sessions occurred in the following general education classrooms: (a) Kate, who was included in a sophomore Algebra class, (b) Ben, who was included in a 10th-12th grade Computer Applications class, and (c) Jacob, who was included in a 11th grade Biology class.

**Materials and Equipment**

Materials needed to implement the study included (a) 4 x 6-in index cards with target words printed on them, (b) science and math worksheets, (c) calculator for math, (d) data sheets, and (e) task analyses. The math worksheet consisted of two linear equations and enough space under each for the participant to write out the steps to solve the problem. The science worksheet consisted of two blank Punnett Squares and two different genetic combinations (e.g., BB x Bb). Worksheets for the writing/using technology chained task were not used, as the task was completed on the computer.

**Data Collection**

The special education teacher collected data on two different sets of skills for each student; each set consisted of one chained task and one discrete task. She recorded correct completion of a response with a “+” and incorrect completion of a response with a “−.” Incorrect responses for chained tasks included (a) performing a step incorrectly, (b) performing a step in an incorrect order, (c) not initiating a step within 5 s of the previous step, (d) not completing a step within 20 s, or (e) making no response. Incorrect responses for discrete tasks included (a) stating the wrong word or shape, (b) not stating a word or shape within 5 s of the task direction, or (c) making no response.

<table>
<thead>
<tr>
<th>Name</th>
<th>Discrete Skill</th>
<th>Chained Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kate</td>
<td>Identify Geometric Figures</td>
<td>Solve Linear Equation in One Variable</td>
</tr>
<tr>
<td></td>
<td>1. Cube</td>
<td>1. Write + or = on each side of equation</td>
</tr>
<tr>
<td></td>
<td>2. Octagon</td>
<td>2. Perform operation</td>
</tr>
<tr>
<td></td>
<td>3. Sphere</td>
<td>3. Write new equation</td>
</tr>
<tr>
<td></td>
<td>4. Cone</td>
<td>4. Write “divide” by multiplier on each side</td>
</tr>
<tr>
<td></td>
<td>5. Pyramid</td>
<td>5. Perform operation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6. Write/state what x equals</td>
</tr>
<tr>
<td>Ben</td>
<td>Using Technology to Publish</td>
<td>Research a Topic on the Internet</td>
</tr>
<tr>
<td></td>
<td>1. Blog – online journal</td>
<td>1. Highlight web address bar</td>
</tr>
<tr>
<td></td>
<td>2. Wiki – website where people can work together</td>
<td>2. Type <a href="http://www.google.com">www.google.com</a></td>
</tr>
<tr>
<td></td>
<td>to edit or change piece of writing</td>
<td>3. Click on google search bar</td>
</tr>
<tr>
<td></td>
<td>3. Scholarly Journal – peer-reviewed periodical</td>
<td>4. Type search word</td>
</tr>
<tr>
<td></td>
<td>containing research</td>
<td>5. Press “search or push “enter” key</td>
</tr>
<tr>
<td></td>
<td>4. Chatroom- online conference site</td>
<td>6. Click on website</td>
</tr>
<tr>
<td></td>
<td>5. Facebook – social networking site</td>
<td></td>
</tr>
<tr>
<td>Jacob</td>
<td>Identify Meaning of Vocabulary Words</td>
<td>Fill Out Punnett Square</td>
</tr>
<tr>
<td></td>
<td>1. Carrying capacity – maximum population size</td>
<td>1. Write letters on top of square</td>
</tr>
<tr>
<td></td>
<td>that environment can sustain</td>
<td>2. Write letters on side of square</td>
</tr>
<tr>
<td></td>
<td>2. Resources – something that satisfies needs of</td>
<td>3. Write letter combination in first block</td>
</tr>
<tr>
<td></td>
<td>living organism</td>
<td>4. Write letter combination in second block</td>
</tr>
<tr>
<td></td>
<td>3. Limiting factors – something that causes</td>
<td>5. Write letter combination in third block</td>
</tr>
<tr>
<td></td>
<td>population to decrease in size</td>
<td>6. Write letter combination in fourth block</td>
</tr>
<tr>
<td></td>
<td>4. Ecosystems – community of living and nonliving</td>
<td></td>
</tr>
<tr>
<td></td>
<td>things interacting as system</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5. Species – group of organisms</td>
<td></td>
</tr>
</tbody>
</table>

**TABLE 1**

Table 1: Discrete and Chained Target Skills for Each Student Participant
**Procedure**

**Simultaneous prompting training.** Prior to beginning instruction, the paraprofessionals and peer tutor received training in the SP procedure from the special education teacher. The 45-min training consisted of an instructional PowerPoint presentation, followed by modeling and individual practice. The PowerPoint included information on the following: (a) systematic instruction, (b) prompting procedures, (c) discrete vs. chained tasks, (d) the target skills for the study, (e) attentional cues and responses, (f) baseline, (g) defining class downtime, (h) giving a task direction, (i) the SP procedure, (j) reliability data, and (k) the students’ working folders. Information on data collection procedures for baseline and test probes was not included since the paraprofessionals and peer tutor would not be delivering any probes. The paraprofessionals and the peer tutor practiced the SP procedure with each other until they reached 100% accuracy for one session, prior to beginning intervention.

**Baseline procedures.** Baseline condition occurred prior to implementing intervention and lasted for a minimum of 3 days. Sessions occurred in a 1:1 instructional format. Each session consisted of one trial for chained tasks and one trial for each of the discrete tasks. After conducting baseline probe sessions at the beginning of the study for each of the participants, the special education teacher continued conducting baseline probe sessions 1 day per week for each participant and skill not currently in intervention.

The special education teacher sat at a table across from the student. For chained tasks, she gave the task direction and handed the student the worksheet to begin. The teacher read the problem on the worksheet aloud to the student and for the math worksheet she gave the student a calculator to use. She gave the student 5 s to initiate each step and 20 s to complete each step. If the student performed an error, the teacher ended the session (i.e., single opportunity format; Cooper, Heron, & Heward, 2007). She then praised the student for on-task behavior. For discrete tasks, the teacher showed an index card with either a vocabulary word or a picture of a geometric shape. She then provided the task direction and gave the student 5 s to respond. The teacher praised all correct responses and ignored all incorrect responses or failures to respond.

**Instructional procedures.** All sessions took place 5 days per week. In the SP procedure, daily probe sessions occur prior to daily training sessions. Daily probe session consisted of the same procedures as baseline probe sessions.

A peer tutor or a paraprofessional conducted daily training sessions. Before each student went to the assigned general education class, the special education teacher gave the student a folder with his or her work for the day (i.e., worksheet for chained task, index cards for discrete task) and the task analysis and data sheets for the instructor. In the general education class, the instructor sat next to the student. Kate sat at a desk in the front row of her math class. Ben sat in his regular seat in a computer applications class, and Jacob sat at a table in the front of his biology class. Instruction occurred during class; (a) when the teacher took attendance, (b) during the morning news program, (c) when students started on their homework, or (d) when students did independent seatwork. Although all of the participants participated in the daily instructional activities of the class, supplemental instruction from the peer tutor or paraprofessional was added during these times. To minimize disruptions, instruction for the investigation did not occur while the general education teacher was teaching a lesson or lecturing the class.

During training sessions for both the discrete and the chained skills, the student performed each skill twice, following the same procedure for both types of skills. For example, in teaching the discrete skill, the instructor first delivered an attentional cue (e.g., “Are you ready to begin working on Math?”) and waited for the student’s attentional response (e.g., “Yes, I’m ready to begin.”). Next, the instructor delivered the task direction immediately followed by the controlling prompt (i.e., verbal model). For example, when showing a picture of a cube, the instructor might say, “What shape is this? It’s a cube.” The instructor praised all correct responses and corrected all errors.

The investigators staggered instruction...
across students. Kate began instruction first. When she reached one day of 100% correct responses on her discrete task, training sessions began for Ben on both his discrete and chained tasks. When he reached one day of 100% correct responses on his discrete task, training sessions began for Jacob on both his discrete and chained tasks. Training sessions continued to a criterion of 100% during three consecutive probe sessions. After one day of criterion, the instructor faded reinforcement from a continuous reinforcement schedule to a VR3 (variable ratio of 3) schedule.

**Maintenance.** The special education teacher conducted maintenance probe sessions in the same manner as baseline and daily probe sessions 1 week and 3 weeks after each student reached criterion for their discrete task and for their chained task. She conducted another maintenance probe session 1 month after all 3 students met criterion on both skills.

**Generalization procedures.** The special education teacher conducted two generalization probe sessions after each student reached criterion on both of his or her target skills. She designed each probe session to assess generalization across persons, settings, and materials. While training sessions occurred in a general education classroom with a peer tutor or paraprofessional using index cards and worksheets, generalization probe sessions occurred in the special education classroom with the special education teacher using materials resembling those used in the state’s AA-AAS and the same format as required for the AA-AAS. During the sessions, the teacher sat across the table from the student with a binder on the table with the generalization probe materials inside. One page with a script and a question faced her. The student had four multiple-choice answers printed on the side facing him or her.

During generalization probe sessions for the geometric shapes, the teacher showed the student a picture of a real-life object (i.e., dice, stop sign, beach ball, party hat, Egyptian ruins) and then asked the student what shape it was. During generalization probe sessions for linear equations, the teacher read a word problem to the student (e.g., “Laura makes $5.00 per hour at her new job. After you type your word or phrase into the Google search bar, what button should you press? How do you finally get the answer to your search question?”). During generalization probe sessions for Punnett Squares, the teacher showed the student an incomplete Punnett Square and had the student choose what went in the blank. During generalization probe sessions for biology terms, the teacher asked for the definition of terms within the context of a story (e.g., “The lifeboat on a cruise ship can hold 50 people. If you are looking up the mascot for Central High School, what should you type in the search bar? After you type your word or phrase into the Google search bar, what button should you press? How do you finally get the answer to your search question?”). During generalization probe sessions for writing, the teacher presented various situations where people want to publish items on the Internet and asked which method would be best for each situation (e.g., “Sally wants to make new friends at her school. She decides to join a social networking site. What is an example of this?”). During generalization probe sessions for searching for information on the Internet, the teacher asked for specific steps in the sequence (e.g., “You are researching the various mascots in high schools across the state. What website can you go to in order to look this up? Where do you find the search criteria? If you are looking up the mascot for Central High School, what should you type in the search bar? After you type your word or phrase into the Google search bar, what button should you press? How do you finally get the answer to your search question?”). During generalization probe sessions for Punnett Squares, the teacher showed the student an incomplete Punnett Square and had the student choose what went in the blank. During generalization probe sessions for biology terms, the teacher asked for the definition of terms within the context of a story (e.g., “The lifeboat on a cruise ship can hold 50 people. If you are looking up the mascot for Central High School, what should you type in the search bar? After you type your word or phrase into the Google search bar, what button should you press? How do you finally get the answer to your search question?”).

**Experimental Design**

Investigators evaluated experimental control using a multiple probe across participants design (MPAP), with concurrent demonstration across two skills per student (Gast & Ledford, 2010). The MPAP was used to minimize student frustration during the baseline phase (over testing) and because students were learning academic skills.

**Social Validity**

To determine effects on the social relationships between the student participants and the peers in their general education class-
rooms, the teacher conducted a survey at the beginning and conclusion of the investigation. She gave the survey to the general education teacher and students in the Algebra class because it was the only class that did not already have students with moderate intellectual disabilities enrolled in it prior to the investigation. The questions on the survey consisted of the following: (1) Do you know anyone with a disability? If yes, how do you know them? (2) Do you think students from the FMD classroom should attend regular classes? Why or why not? (General education students only) or Prior to this year, have you ever taught someone from a FMD classroom? If yes, how many students? When? (General education teachers only); (3) Do you think students from the FMD classroom should be taught STEM content? Why or why not? (4) What benefits do you think students from the FMD classroom could gain from being in regular classes? (social, academic, communication skills, self-esteem, nothing).

Reliability

A graduate student with a degree in special education collected procedural fidelity (PF) and interobserver agreement (IOA) reliability data during probe sessions. Instructor behaviors included (a) delivering attentional cue, (b) waiting for attentional response, (c) presenting stimulus, (d) ending session following error in chained task or ignoring errors during discrete task, (e) recording data, and (f) praising correct responses and good behavior. The graduate student collected PF and IOA reliability data during 33%, 25%, and 20% of baseline sessions for Kate, Ben, and Jacob, respectively. She collected PF and IOA reliability data during 50% of discrete task test probes for Kate and Ben and 29% for Jacob. She collected reliability data for 30%, 38% and 40% of chained task test probes for Kate, Ben and Jacob, respectively. She collected PF and IOA reliability during 33% of maintenance probe sessions and 50% of generalization probe sessions for all students.

The special education teacher collected PF data during training sessions. During training sessions, instructor behaviors included (a) waiting until it was appropriate to begin training, (b) delivering attentional cue, (c) presenting stimulus, (d) immediately delivering controlling prompt, (e) praising correct responses or correcting errors, and (f) performing task twice. The special education teacher collected reliability data during 33% of Kate’s discrete task training sessions and 20% of her chained task training sessions. She collected PF reliability data during 25% of Ben’s training sessions. She collected PF reliability during 43% of Jacob’s discrete task training sessions and 60% of his chained task training sessions.

Another FMD teacher collected PF data during the SP procedure training session conducted by the special education teacher for the paraprofessional and the peer tutor. Teacher behaviors included (a) showing instructional PowerPoint, (b) modeling SP procedure, (c) allowing instructors to practice SP procedure with each other, (d) collecting data on ability to implement SP procedure with 100% accuracy, and (e) reviewing student folders with each instructor.

The investigators calculated PF agreement by dividing the number of completed steps by the number of possible steps and multiplying by 100 (Billingsley, White, & Munson, 1980). They calculated IOA by dividing the number of agreements by the number of agreements plus disagreements and multiplying by 100 (Gast, 2010). PF reliability agreement was 100% for all probe sessions across students. IOA reliability agreement was also 100% during baseline probe and daily probe sessions across students. PF reliability agreement was 100% during Kate’s discrete task training sessions and ranged from 83% to 100% (mean = 92%) during her chained task training sessions. During one training session, the peer tutor forgot to praise Kate for completing a step correctly. PF reliability agreement was 83% during Ben’s discrete task training sessions (collected once because he reached criterion after two sessions). The paraprofessional did not deliver the controlling prompt before Ben said one of his vocabulary words during one session. During Ben’s chained task training sessions, PF agreement ranged from 83% to 100% (mean = 92%). During one session, the paraprofessional did not deliver the controlling prompt before Ben pushed the “search” button on the computer. PF agreement during Jacob’s discrete task train-
ing sessions was 100%. During chained task training sessions, PF agreement ranged from 83% to 100% (mean = 92%). Jacob filled in one of the boxes of the Punnett square before the paraprofessional delivered the controlling prompt during one session. PF agreement for the SP training for instructors conducted by the teacher was 100%.

Results

The SP procedure was effective in teaching STEM content skills to three secondary students with moderate intellectual disabilities within general education classroom settings. The effectiveness of the SP procedure for the three students can be viewed in Figure 1. In addition, the paraprofessionals and the peer tutor implemented the intervention with a high degree of procedural fidelity.

Effectiveness and Efficiency Data

During baseline sessions, all three students’ data were stable at 0% for both the discrete task and chained task. Kate met criterion for her discrete task in six instructional sessions and for her chained task in 10 instructional sessions. The total time for discrete task instructional sessions was 12 min (mean = 2 min per session). The total time for chained task instructional sessions was 56 min (mean = 5 min 36 s per session). Ben met criterion for his discrete task in four sessions and for his chained task in eight sessions. His total time for instruction on the discrete task was 9 min (mean = 2 min 15 s per session) and 43 min for his chained task (mean = 5 min 24 s per session). Jacob met criterion for his discrete task in seven sessions, totaling 20 min (mean = 3 min per session). He met criterion for his chained task in five instructional sessions, totaling 14 min (mean = 2 min 48 s per session).

Maintenance and Generalization

All three students maintained their target skills at 100% criterion for 1 month after intervention ended. The students also generalized at least some of their target skills during probe sessions resembling the state’s AA-AAS: (a) Kate - mean of 100% correct responses for discrete and 80% for chained task, (b) Ben - mean 100% correct responses for discrete task and 60% for chained task; and (c) Jacob - mean of 60% correct responses for discrete task and 100% for chained task.

Social Validity

Results of the social validity survey were, for the most part, positive. Prior to the beginning of the investigation and Kate’s inclusion in the Algebra class, 12 of the typically developing students said that they thought students with moderate intellectual disabilities should attend regular classes and five thought they should not; after the study, 15 students said students with MSD should attend. All 17 students thought students with moderate intellectual disabilities could learn core content following the investigation. Prior to the investigation, 16 of the peers thought that students with moderate intellectual disabilities should be taught core content, and only one did not. Following the investigation, all 17 students thought students with moderate intellectual disabilities should be taught core content. Prior to the investigation, the general education students identified several benefits of students with moderate intellectual disabilities being included in the regular classroom that increased following the investigation: (a) social interaction (increase from 8 to 10), (b) academic gains (increase from 5 to 7), (c) communication skills (remained the same at 8), and (d) self-esteem (increase from 6 to 9).

The Algebra teacher had mostly positive responses prior to the investigation. He stated that he thought students from the FMD classroom should attend regular classes, could learn core content, and would benefit from inclusion in the areas of social integration and academic advantages. He stated that he did not think students with moderate intellectual disabilities should be taught core content, as pacing would be a serious issue. After the investigation, the same teacher said that he thought students from the FMD classroom should be taught grade level core content, just at a modified pace.

Discussion

The purpose of the current investigation was to evaluate the effectiveness of an embedded
SP procedure in teaching STEM content and skills to students with moderate intellectual disabilities in an inclusive setting. All students mastered their target skills after a peer tutor or a paraprofessional provided an embedded SP procedure within the normal routines in a general education classroom. Students also maintained their target skills for at least 1 month following instruction and generalized the skills with 60% to 100% correct responding.
One challenge to including students with moderate intellectual disabilities in general education classes is that the pace of instruction is usually much quicker than in special education classes. This concern is especially valid when students are learning more challenging content or chained skills in higher grades. The use of embedded SP in the current study, however, may negate these concerns since the student participants reached 100% on their discrete tasks in an average of 3.7 sessions and 100% on their chained tasks in an average of five sessions. Similarly, Reisen et al. (2003) compared SP to CTD to teach discrete tasks and found that three of the four participants reached 100% in just three sessions.

Other studies also have demonstrated that the SP procedure can be effective in teaching core content skills to students with disabilities. For example, Head et al. (2011) and Pennington, Stenhoff, Gibson, and Ballou (2012) both used SP to teach core content to students with high incidence disabilities. The current investigation extended this research and the research conducted by Collins et al. (2007) with students with moderate intellectual disabilities by using an embedded SP procedure to teach STEM content to students with moderate intellectual disabilities in inclusive classrooms. Embedded SP by a peer, paraprofessional, or the classroom teacher may benefit a diverse group of students (e.g., students with learning disabilities, English language learners, students at risk) in learning STEM content in inclusive contexts.

In addition to the current study, generalization and maintenance of learned skills also has been a concern in previous studies teaching core content to students with moderate intellectual disabilities. For example, in a study by Knight, Smith, Spooner, and Browder (2011), students learned science descriptors using objects during the intervention condition and were able to generalize to novel objects but had more difficulty generalizing when using pictures or within an inquiry-based lesson. In the current study, the special education teacher planned with the general education teacher to carefully select STEM skills that a student would likely use in the future in various settings or that would be considered foundational knowledge for more complex material.

The use of a peer tutor to implement instruction was an important component of the intervention. Peers and paraprofessionals found the SP procedure easy to use and made few errors, leading to a high degree of procedural fidelity. Previous studies have supported these findings when using systematic instructional procedures. For example, Jimenez et al. (2012) showed that peer tutors could implement the CTD procedure within an inclusive science class to teach science vocabulary words, pictures, and concept statements to students with moderate intellectual disabilities, while Collins, Branson, Hall, and Rankin. (2001) showed that a peer tutor could implement the SLP procedure within an inclusive English class to teach writing. One of the contributions of this study is that a peer tutor implemented the SP procedure with high fidelity in teaching both a discrete task and a chained task in an inclusive math classroom.

Working with peers may have led to students experiencing social benefits. The pre-and post-intervention social validity surveys completed by the students in the general education Algebra class indicated increased acceptance of the students with moderate intellectual disabilities by the general education students. One of the students originally stated that the students from the FMD classroom should not be allowed to attend regular classes because “. . . people will make fun of them and cause them emotional pain. As much as I want to see them have a chance to see others, it is my experience that people are cruel.” Following the investigation, this same student said that students from the self-contained classroom should be allowed to attend regular classes because “. . . from what I’ve seen in here, no one made her feel uncomfortable or sad.” This student had indicated that he had a cousin and several friends with disabilities. After the investigation, he wrote that he had a disability as well. It is possible that his past personal experiences had influenced his initial apprehension, but, after having Kate included in his class, he was able to see that peers could be accepting of differences.
Implications for Practitioners

With the increased focus in public schools on STEM education, it is important for teachers to consider ways in which all students can participate and gain generalizable skills in this content. In the current study, special and general education teachers collaborated to determine targeted skills that linked to the standards and that were reasonable to teach in a short period of time. More importantly, generalizable, maintainable skills were selected with the intent to benefit students with disabilities now and in the future, as these skills were based on students’ current interests and post secondary goals. Although the area STEM may be intimidating for some special education teachers, the current study demonstrates how students with moderate disabilities can learn individualized grade-appropriate STEM content linked to post secondary goals in an inclusive classroom. While the pace of instruction in general education settings (especially in high school) is certainly a barrier for many students with disabilities, use of the SP may be one way to combat this barrier since it is a brief, yet effective, instructional strategy. In addition, this study illustrates how social benefits can be a collateral gain when students with disabilities are included with their same-aged peers in STEM.

Limitations and Future Research

Several limitations should be noted when considering the results of this investigation. First, a small number of students participated in the investigation, limiting the validity of the findings. Only two investigations to date (Collins, 2007; Reisen et al., 2003) have examined the effectiveness of the SP procedure in teaching core content to students with MSD in inclusive settings, and there were only four participants in each of those investigations. Additional research is needed before this can be considered an evidence-based practice.

A second limitation is the skills that were taught. Each student received instruction on five vocabulary words or geometric shapes and one chained task. The researchers determined that these skills were equal in complexity, but the results of the study may indicate otherwise. Ben and Jacob made immediate progress with their chained tasks and reached 100% after only four and three sessions, respectively. Kate, on the other hand, did not make any progress during the first four sessions and took eight sessions to reach 100% criterion. This could be due to the fact that, even though all chained tasks had six steps in the task analysis, solving linear equations may be a more complex (or less motivating) task than filling out Punnett squares or looking up information on the Internet.

A third limitation was the use of single opportunity probes for assessing chained tasks during probe sessions. Baseline levels may not have been so low if the students had been given the opportunity to complete each step of the task analysis rather than being stopped after their first error, which occurred on the first step for all of the participants. The researchers used single opportunity probes to reduce the likelihood that learning would occur if the instructor completed some of the steps during probe conditions, but this may not have accurately reflected the ability of the students to perform some of the steps.

A fourth limitation was in the timing that the skills were taught. Instruction took place when other students were doing independent seatwork or other similar situations in the general education classroom in order to prevent distractions to the other students; however, it may not be considered a truly “inclusive” time to teach the skill. Future researchers should require more involvement from general education teachers in the delivery of systematic instruction. General educators or peers could implement the SP procedure during typical instructional times. If general educators were actively involved in systematic interventions, it could enlighten them on the benefits of inclusion and the practical inclusion strategies for instruction, as indicated by the survey response of the Algebra teacher in this investigation.

The fifth limitation was in the manner in which probe sessions were conducted. Probes were conducted in the special education classroom by the special education teacher. This limited distractions and ensured greater reliability of the data, as the special education teacher was highly trained in the SP procedure and data collection procedures. However, future research should look at data col-
lection in a natural setting (i.e. the inclusive classroom).

Finally, the use of peer tutors is more socially acceptable than the use of a paraprofessional in inclusive classrooms. Carter, Sisco, Melekoglu, and Kurkowski (2007) showed that students with disabilities engaged in more social interactions when working with a peer than when working with a paraprofessional. Due to scheduling conflicts, both a peer and a paraprofessional delivered instruction in the current study, demonstrating that participants could acquire the academic skills from both types of instructors in similar time frames. Future research, however, should focus on the use of peers in inclusive settings and determine if peer-directed instruction in inclusive classrooms will result in both the acquisition of academic skills and increased social interactions. Additional research could examine methods aimed at increasing both academic and social benefits in the context of real-world, problem-based learning (e.g., engage in inquiry process using cooperative small groups of typically-developing peers to learn processes of science as well as communication, such as asking questions).

One of the goals of this investigation was to teach STEM related content and skills to students eligible for the AA-AAS. Another suggestion for future research would be to compare the AA-AAS scores of students taught in segregated classroom to the scores of students taught with systematic instruction in inclusive settings since this could provide further support for the inclusion of students with MSD.

Conclusion

As STEM continues to receive national attention and additional states continue to adopt CCSS and NGSS, educators need effective and efficient instructional strategies for supporting students with a wide range of needs in inclusive classrooms. Results from this study support the use of SP by instructors other than the special or general education teacher for teaching STEM content to students with complex needs. Further, this study shows that educators can work together to benefit all of the students in their classroom, both socially and academically. Negative perceptions can be changed as a result of increased access to natural contexts and general education curricula, as evidenced by the change in attitudes of the peers and classroom teacher. Since embedded SP (a) is easy to implement, (b) is less time consuming than many other systematic instructional procedures, (c) can result in fewer errors than other prompting strategies, and (d) can be used by a variety of instructors, including teachers, paraprofessionals, and peers with a high degree of fidelity, educators may want to consider using practice in inclusive contexts to promote progress in, maintenance, and generalization of STEM content areas.

References


Received: 19 November 2014
Initial Acceptance: 17 January 2015
Final Acceptance: 4 May 2015

An Evaluation of Constant Time Delay and Simultaneous Prompting Procedures in Skill Acquisition for Young Children with Autism

Julie A. Ackerlund Brandt, Sara Weinkauf, Nicole Zeug, and Kevin P. Klatt
University of Wisconsin-Eau Claire

Abstract: Previous research has shown that various prompting procedures are effective in teaching skills to children and adults with developmental disabilities. Simultaneous prompting includes proving a prompt immediately following an instruction; whereas constant time-delay procedures include a set time delay (i.e., 5 s or 10 s) prior to delivering a prompt following an instruction. These prompting procedures have been previously compared with mixed results. The current study used an alternating treatments design to compare simultaneous prompting to a constant time-delay procedure to evaluate efficacy and efficiency of each procedure, in addition to the number of errors which occurred under each condition. Results from the current study are discussed as well as limitations and future directions.

Time (prompt) delay procedures (i.e., constant and progressive time delay), involve initially presenting an instruction and immediately prompting the correct response. After a few trials of providing an immediate prompt, a short delay occurs between the instruction and the prompt. This delay may either remain a constant duration of time (i.e., constant time delay) or an incremental increase in duration on consecutive trials (i.e., progressive time delay). The purpose of this delay is to provide an opportunity for the child to respond independently (thereby transferring control of the response from the prompt to the instruction). The constant time delay has been used to teach a variety of discrete and chained behaviors, and has been identified as an evidenced-based procedure for teaching sight words and picture recognition (Browder, Ahlgrim-Delzell, Spooner, Mims, & Baker, 2009). Many variations of a constant time-delay procedure have been used to teach vocal/verbal language skills (Carbone, Sweeney-Kerwin, Attanasio, & Kasper, 2010; Charlop, Schreibman, & Thibodeau, 1985; Halle, Marshall, & Spradlin, 1979; Ingenmy & Van Houten, 1991), academic (Cates et al., 2007; Coleman-Martin & Heller, 2004; Heal, Hanley, & Layer, 2009), self-help (McDonnell & Ferguson, 1989), and leisure skills (Wall & Gast, 1997).

In addition to the constant time delay, simultaneous prompting has been used successfully to transfer stimulus control. This procedure involves presenting a prompt immediately following the instruction on all trials. There is no opportunity to respond independently, and therefore, probe trials are conducted each day of instruction to determine whether stimulus control has been transferred. Simultaneous prompting has been used effectively to teach a variety of academic tasks (Akamanoglu & Batu, 2004, Akamanoglu-Uludag & Batu, 2005, Leaf, Sheldon, & Sherman, 2010; Riesen, McDonnell, Johnson, Polychronis, & Jameson, 2003), and self-help skills (Kurt & Tekin-Iftar, 2008; Parrott, Schuster, Collins, & Gassaway, 2000; Sewell, Collins, Hemmeter, & Schuster, 1998).

This study was completed as the first author’s undergraduate capstone project for the Ronald E. McNair Program at the University of Wisconsin – Eau Claire. The authors would like to acknowledge and thank the other undergraduate research assistants who helped in data collection for this study: Sara Czekalski, Sara Tillman, and Britta Fiksdal. Correspondence concerning this article should be addressed to Kevin P. Klatt, Psychology Department, University of Wisconsin-Eau Claire, Eau Claire, WI 54702. E-mail: klattkp@uwec.edu

This study was completed as the first author’s undergraduate capstone project for the Ronald E. McNair Program at the University of Wisconsin – Eau Claire. The authors would like to acknowledge and thank the other undergraduate research assistants who helped in data collection for this study: Sara Czekalski, Sara Tillman, and Britta Fiksdal. Correspondence concerning this article should be addressed to Kevin P. Klatt, Psychology Department, University of Wisconsin-Eau Claire, Eau Claire, WI 54702. E-mail: klattkp@uwec.edu

Correspondence concerning this article should be addressed to Kevin P. Klatt, Psychology Department, University of Wisconsin-Eau Claire, Eau Claire, WI 54702. E-mail: klattkp@uwec.edu

This study was completed as the first author’s undergraduate capstone project for the Ronald E. McNair Program at the University of Wisconsin – Eau Claire. The authors would like to acknowledge and thank the other undergraduate research assistants who helped in data collection for this study: Sara Czekalski, Sara Tillman, and Britta Fiksdal. Correspondence concerning this article should be addressed to Kevin P. Klatt, Psychology Department, University of Wisconsin-Eau Claire, Eau Claire, WI 54702. E-mail: klattkp@uwec.edu
time delay and simultaneous prompt procedures (Head, Collins, Schuster, & Ault, 2011; Kurt and Tekin-Iftar, 2008; Riesen et al., 2003; Schuster, Griffent, & Wolery, 1992; Seward, Schuster, Ault, Collins, & Hall, 2014; Tekin and Kircaali-Iftar, 2002). Both prompting procedures have been shown to be effective for teaching a variety of skills. For example, Schuster et al. (1992) taught sight words to elementary students diagnosed with intellectual disabilities. Although both procedures were effective in teaching sight words, the simultaneous prompting procedure required fewer trials, sessions, and training time to meet the pre-established criterion. In addition, the simultaneous prompting procedure resulted in fewer errors during teaching sessions and daily probes.

Additionally, Tekin and Kircaali-Iftar (2002) taught three children with intellectual disabilities to receptively identify animals, and both procedures were, again, equally effective. Kurt and Tekin-Iftar (2008) taught leisure skills to four children diagnosed with autism. The results showed that both procedures were equally effective in teaching leisure skills for three of the four children. Two of the children also had fewer errors with the constant time delay and the other two had fewer errors with the simultaneous prompting procedure.

Head et al. (2011) compared the procedures in teaching state capitals to four high school students with learning and behavior disorders. The results showed both procedures were effective in teaching the state capitals. There was no clear difference across students regarding the number of errors, although individual differences showed slight advantage to one or the other procedure, depending on the participant.

Although the majority of research has shown that both procedures are equally effective, there are instances where one procedure may be more effective than the other. Riesen et al. (2003) also taught academic skills (i.e., read or define vocabulary words) to four students diagnosed with moderate to severe intellectual disabilities. Results showed that, although both procedures were effective in teaching academic skills, the constant time delay procedure was slightly more effective for two participants, and the simultaneous prompting procedure was more effective for the other two participants. In a more-recent study, Seward et al. (2014) taught two solitaire card games to five high school students with moderate intellectual disabilities. The results showed both procedures were effective with four students; however, the simultaneous prompting procedure was more effective for one student. Additionally, there were fewer errors during probes in the constant time delay, and less errors in teaching with the simultaneous prompting procedure.

In general, it is important to note that both prompting procedures seem to be effective teaching procedures for students with intellectual disabilities; however, the efficiency of the two procedures has also been evaluated. For example, Tekin and Kircaali-Iftar (2002) also measured the amount of training time necessary across procedures. The results showed that the simultaneous prompting procedure resulted in less training time than the constant time delay prompting procedure. This is another important variable to be considered when evaluating these two prompting procedures.

The results from these studies suggest at least two findings. First, both procedures seem to be effective in teaching a variety of skills, with occasional individual differences in effectiveness for one or the other procedure. Second, results from the studies suggest differences in efficiency (e.g., number of errors). Overall, fewer errors were found using the simultaneous prompting procedure in two studies (Schuster et al., 1992; Tekin & Kircaali-Iftar, 2002), errors were not measured in one study (Riesen et al., 2003), and mixed results were found in the most recent studies (Kurt & Tekin-Iftar, 2008; Head et al., 2011; Seward et al., 2014). The measurement of errors, however, can be further divided between those that occur in daily probes (required only in simultaneous condition) versus training sessions. That is, errors potentially can occur when the skill is probed each day, and also during training (although this is not likely with simultaneous prompting since the correct answer is immediately prompted). In the studies conducted to date, results pertaining to errors differed across studies. These differences include simultaneous prompting resulting in fewer errors during both probes and training sessions (Schuster et al.), fewer errors.
in probes for constant time delay and fewer errors in training sessions for simultaneous prompting (Seward et al.), errors were not recorded (Riesen, et al.), combined across probes and training (Kurt & Tekin-Iftar), reported only in probes because none occurred in training (Head et al.), or procedures were not clear whether errors were combined or only from training (Tekin & Kircaali-Iftar). Therefore, drawing conclusions about the differences between the two procedures regarding errors would be premature at this point.

Thus far, data from previous research shows both prompting procedures effective in teaching skills and neither prompting procedure clearly more efficient. No studies to date, however, have compared these two procedures for young children diagnosed with autism under the age of six. The purpose of this study, therefore, was to compare the constant time delay and simultaneous prompting procedures with young children with autism, and to specifically investigate the differences in errors in both daily probes and teaching sessions.

Method

Participants

Participants included one female and six males diagnosed with autism ranging in age from two to six years old. Five of the participants were receiving behavioral intervention services for approximately 30–35 hours a week in their homes. Two participants were receiving behavioral intervention services four hours a week at a university-based autism clinic.

Therapists included the undergraduate students at a local university. All therapists had completed at least one semester of intensive internship training in applying behavioral principles to teaching children diagnosed with autism. The lead therapists, the first three authors, met with other therapists on their teams prior to beginning a program to ensure that each program was conducted correctly and with high treatment fidelity.

Settings and Materials

The study was conducted in a therapy room at a university-based autism clinic for two participants and in a designated area of the home for the other five participants. Participants and therapists sat across from each other at a small table or on the floor. Preferred toys, food, and beverages were used as putative reinforcers for correct responding during sessions. Various materials such as pictures and flashcards were used to implement the various programs.

Data Collection and Response Measurement

The primary dependent variables were the number of targets to mastery criterion and number of errors during teaching and probe sessions. During sessions, data were collected on correct responses, no responses, and errors emitted during trials. A correct response was defined as an accurate, independent response within 5 s of instruction. A no response was defined as the omission of a response within 5 s of the instruction. An error was defined as an incorrect response within 5 s of the instruction. The mastery criterion for each target was 100% correct during a daily probe session.

Interobserver Agreement and Treatment Integrity

Interobserver agreement (IOA) of the dependent variables was collected during 9%–100% of sessions for each participant. Interobserver agreement was calculated by comparing each observer’s data using a trial-by-trial method. Agreements for each trial were defined as each observer scoring the same responses. The trials scored as agreements were summed and divided by the total number of trials and multiplied by 100%. IOA was calculated for 18% of sessions with Shawn and ranged from 95–100%. IOA was calculated for 100% of sessions with Max and ranged from 97–100%. IOA was calculated for 15% of sessions with Nick and ranged from 88–100%. IOA was calculated for 9% of sessions with Brian and was 100% across all sessions. IOA was calculated for 77% of sessions with Maria and ranged from 94–100%. IOA was calculated for 88% of sessions with Derek and ranged from 97–100%. IOA was calculated for 33% of sessions with Edward and was 100% across all sessions.

Treatment integrity was measured using a
procedural task analysis and was calculated by summing the number of steps implemented correctly by the therapist, dividing by the total number of steps and multiplying by 100%. Treatment integrity was collected for 6%–18% of sessions for each participant and was 100%.

Experimental Design

An adapted alternating treatments design was used to compare the simultaneous prompting and constant-time delay procedures (Sindelar, Rosenberg, & Wilson, 1985).

Procedure

Specific target skills for each participant were identified from their individualized treatment plan. Several targets (two to four per condition, total of four to eight targets) for each specific skill were identified and randomly assigned to one of the two conditions. For example, if the specific skill was addition, three targets (e.g., 1+2, 3+4, 5+6) were assigned to the simultaneous prompt condition and three more targets (e.g., 2+4, 3+5, 1+6) were assigned to the constant prompt-delay condition for a total of six targets. Considerations were taken to ensure that targets were relatively

<table>
<thead>
<tr>
<th>Participant</th>
<th>Age</th>
<th>Target Skill</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shawn</td>
<td>6</td>
<td>Expressive Addition</td>
</tr>
<tr>
<td>Max</td>
<td>2</td>
<td>Matching Associate Pictures</td>
</tr>
<tr>
<td>Nick</td>
<td>4</td>
<td>Expressive Adverbs</td>
</tr>
<tr>
<td>Brian</td>
<td>4</td>
<td>Expressive Adverbs</td>
</tr>
<tr>
<td>Maria</td>
<td>3</td>
<td>Receptive Picture Identification</td>
</tr>
<tr>
<td>Derek</td>
<td>2</td>
<td>Receptive Object Identification</td>
</tr>
<tr>
<td>Edward</td>
<td>3</td>
<td>Expressive Adjectives</td>
</tr>
</tbody>
</table>
equal in complexity and difficulty. For example, when math problems were the target, all problem included single digits and were all either addition or subtraction. If the response was a vocal/verbal response, the responses were matched such that approximately the same number of words was required across targets. Table 1 includes a summary of the target skills for each participant.

Probe trials. At the beginning of each session, three probe trials were conducted for each target which was still in acquisition. Depending on the number of current targets, there were a total of three to 24 probe trials. For each probe trial, the therapist presented the instruction and waited up to 10 s for the participant to respond. Correct responses resulted in praise and there were no programmed consequences for errors. Following a response, or if 10 s elapsed without a response, the therapist presented the instruction for the next trial. If a target was mastered, it was removed and only the remaining targets were taught during the following session.

Teaching trials. Following the probe trials, targets which were not mastered were taught using either the simultaneous or constant time delay prompting procedures. Each target was presented five times, for a total of five to 40 teaching trials, depending on the number of non-mastered targets. All of the targets in both the simultaneous prompting and constant time delay conditions were taught in a block of five to 20 trials. The targets within each block were presented in a random order and the condition blocks were counterbalanced across sessions. Each teaching trial consisted of the presentation of an instruction, implementation of one of

Figure 2. The top panel depicts the cumulative number of target responses that reached criterion in the constant time delay and simultaneous probe sessions. The bottom panel depicts the total number of errors during teaching and probe sessions across simultaneous and constant time delay conditions.
the prompting procedures (see below), and the delivery of a putative reinforcer based on responding.

**Simultaneous prompting.** During simultaneous prompting trials, an immediate prompt of the correct answer (e.g., “What is this? Say ‘cat.’”) was delivered after the instruction on all trials. Correct responses in this condition resulted in praise and the delivery of an edible or tangible item (e.g., a token) and the presentation of the next instruction. Incorrect or no responses resulted in no programmed consequences and the presentation of the next trial.

**Constant time delay.** During constant-time delay trials, an immediate prompt (e.g., “What is this? Say ‘dog.’”) was delivered after the instruction on the first trial of each session per target. During the remaining four trials, a 4 s delay before the prompt (e.g., “What is this?” pause for 4 s., “Say ‘bird.’”) was implemented. Correct responses in this condition resulted in praise and the delivery of an edible or tangible item (e.g., a token) and the presentation of the next instruction. Incorrect or no responses to the first trial in this condition resulted in no programmed consequences and the presentation of the next trial. Incorrect responses on the second to fifth trials resulted in the therapist providing corrective feedback and error correction. Error correction was implemented by saying “This is a bird, say bird” and re-presenting the instruction.

**Results**

Results for each participant are shown in Figures 1–7. The top panel in each figure depicts the cumulative number of target responses that reached criterion (100% correct during a
daily probe session) in both conditions in each session. The bottom panel depicts the total number of errors in probe and teaching sessions for both conditions.

Results for Max are shown in Figure 1. Three targets were taught in both the simultaneous and constant time delay conditions. Max reached criterion for all three targets in the simultaneous condition after four sessions, and all three targets in the constant time delay condition after four sessions. In probe sessions, Max had a total of seven errors in the simultaneous condition and three errors in the constant prompt condition. In teaching sessions, Max had a total of zero errors in the simultaneous condition and 10 errors in the constant prompt condition.

Results for Shawn are shown in Figure 2. Three targets were taught in both the simultaneous and constant time delay conditions. Shawn reached criterion for all three targets in the simultaneous condition after eight sessions, and all three targets in the constant time delay condition after five sessions. In probe sessions, Shawn had a total of 24 errors in the simultaneous condition and 31 errors in the constant prompt condition. In teaching sessions, Shawn had a total of zero errors in the simultaneous condition and 18 errors in the constant prompt condition.

Results for Nick are shown in Figure 3. Four targets were taught in both the simultaneous and constant time delay conditions. Nick reached criterion for all four targets in the simultaneous condition after 11 sessions, and all four targets in the constant time delay condition after five sessions. In probe sessions, Nick had a total of 49 errors in the simultaneous condition and 28 errors in the constant prompt condition. In teaching sessions, Nick
had a total of 33 errors in the simultaneous condition and 42 errors in the constant prompt condition.

Results for Brian are shown in Figure 4. Eight targets were taught in both the simultaneous and constant time delay conditions across two sets. In the first set, Brian reached criterion for all five targets in the simultaneous condition after two sessions, and all five targets in the constant time delay condition after five sessions. In the second set, Brian reached criterion for all three targets in the simultaneous condition after four sessions, and all three targets in the constant time delay condition after one session. In probe sessions, Brian had a total of 15 errors in the simultaneous condition and 13 errors in the constant prompt condition. In teaching sessions, Brian had a total of zero errors in the simultaneous condition and 20 errors in the constant prompt condition.

Results for Maria are shown in Figure 5. Two and four targets were taught in both the simultaneous and constant time delay conditions across two sets. In the first set, Maria reached criterion for both targets in the simultaneous condition after four sessions, and both targets in the constant time delay condition after one session. In the second set, Maria reached criterion for all four targets in the simultaneous condition after seven sessions, and all four targets in the constant time delay condition after seven sessions. In probe sessions, Maria had a total of 35 errors in the simultaneous condition and 38 errors in the constant prompt condition. In teaching sessions, Maria had a total of one error in the simultaneous condition and 38 errors in the constant prompt condition.

Figure 5. The top panel depicts the cumulative number of target responses that reached criterion in the constant time delay and simultaneous probe sessions. The bottom panel depicts the total number of errors during teaching and probe sessions across simultaneous and constant time delay conditions.
Results for Derek are shown in Figure 6. Two targets were taught in both the simultaneous and constant time delay conditions. Derek reached criterion for both targets in the simultaneous condition after three sessions, and both targets in the constant time delay condition after six sessions. In probe sessions, Derek had a total of 10 errors in the simultaneous condition and 10 errors in the constant prompt condition. In teaching sessions, Derek had a total of four errors in the simultaneous condition and 18 errors in the constant prompt condition.

Results for Edward are shown in Figure 7. Three targets were taught in both the simultaneous and constant time delay conditions. Edward reached criterion for all three targets in the simultaneous condition after three sessions, and all three targets in the constant time delay condition after three sessions. In probe sessions, Edward had a total of 15 errors in the simultaneous condition and 23 errors in the constant prompt condition. In teaching sessions, Edward had no errors in the simultaneous condition or constant-time delay condition.

**Discussion**

The results from this study replicated and extended previous research in that both the constant time delay and simultaneous prompting procedures can be used to teach skills to young children with autism. Overall, the results also suggested neither the constant time delay nor the simultaneous prompting procedure was more effective. These results are consistent with the previous studies that have
compared these prompting procedures, and extends this literature to young children diagnosed with autism.

In this study, there were individual differences in errors in probe sessions for both prompting procedures. More errors occurred in the simultaneous prompt condition during probe sessions for Max, Nick, and Brian. More errors occurred in the constant time delay condition during probe sessions for Shawn, Maria, and Edward. No difference between conditions was found for Derek. These results, therefore, do not show either procedure to be more efficient in probe trials.

There were consistent differences, however, in errors in teaching sessions. More errors occurred in the constant time delay condition for Max, Shawn, Nick, Brian, Maria, and Derek. No errors occurred in either condition for Edward. This result is consistent with two previous studies showing fewer errors using the simultaneous prompt procedure in teaching skills (Schuster et al., 1992; Sward et al., 2014). More errors likely occurred in the constant time delay condition due to more opportunity for error inherent in the procedure. The time period between the instruction and the response (or prompt) provides an opportunity for a child to err. The simultaneous prompting procedure, on the other hand, includes an immediate prompt and therefore there is less opportunity to err.

**Limitations**

The results from this study should be interpreted with caution for several reasons. First,
the number of targets in each condition changed according to when the criterion was met. When a target met the criterion, it was removed from that condition; therefore, the number of targets taught in each condition could be different from one session to the next. Although there may have been differences from session to session, the two prompting procedures were conducted as they typically have been in past studies (and in applied settings). Second, the duration to implement each procedure was not recorded, and therefore conclusions cannot be made regarding whether differences existed in the duration of time to implement each procedure. Third, the frequency of IOA and treatment integrity checks was lower than the standard (25%) for some participant. This limitation is somewhat mitigated, however, by the consistently high results associated with the IOA (88–100%) and treatment integrity (100%).

Future Research

Future research should investigate maintenance and generalization of skills learned using these two prompting procedures. Previous research has shown that skills taught using constant time delay procedures result in greater maintenance than targets taught using simultaneous procedures, although this research needs to be extended to children with autism.

Future research pertaining to both prompting procedures needs to be conducted to determine whether either procedure is more effective and efficient on an individual level, and if so, how to quickly make that determination. For example, research could be conducted to determine whether children who engage in problem behavior maintained by negative reinforcement (e.g., avoidance/escape from demands) might benefit from the simultaneous prompting procedure because the instructor provides an immediate prompt, thereby potentially making escape less valuable.

Future research could also investigate whether instructors, including teachers and parents working with children in their home setting, have a preference for either prompting procedure. Relatedly, an investigation could be conducted to determine whether children have a preference for either procedure when learning new skills. The results from these possible investigations, along with the current study, are important in determining how best to effectively and efficiently teach skills to children diagnosed with autism.

Summary

The current study demonstrated that both prompting procedures can be used to teach skills to children diagnosed with autism. More errors were found when teaching skills using the constant time delay procedure, and therefore teachers should be aware of using this procedure for children with autism who engage in problem behavior maintained by avoidance or escape from demands.

References


Halle, J. W., Marshall, A. M., & Spradlin, J. E.


Received: 15 January 2015
Initial Acceptance: 23 March 2015
Final Acceptance: 15 June 2015
Family Generated and Delivered Social Story Intervention: Acquisition, Maintenance, and Generalization of Social Skills in Youths with ASD

Seray Olçay-Gül  
Hacettepe University

Elif Tekin-Iftar  
Anadolu University

Abstract: The purpose of this study was to examine whether (a) family members were able to learn to write a social story and deliver social story intervention to teach social skills to their children (age 12 to 16) with ASD, (b) youths with ASD acquired and maintained the targeted social skills and generalized these skills across novel situations. Multiple probe design across three dyads (family members-youths with ASD) was used. Results showed that family members were able to write and deliver social story intervention with high treatment integrity and youths acquired the targeted social skills. Also they were able to maintain the acquired skills over time and generalize them to novel situations. Family members reported positive opinions about using social stories and social comparison data showed that after intervention, the social skill performance of the youth with ASD was found to be similar to their peers. Based on the findings, future research needs and implications for practice are discussed.

Social stories (SS) are defined as the process of sharing structured stories describing specific social situations in which a child with autism will encounter and appropriate responses to the social stimuli that will be encountered in that situation (Gray & Garand, 1993). They should be brief and written with certain formats and rules (Barry & Burlew, 2004; Gray & Garand). The implementation of it requires two main steps: (a) writing a story and (b) implementing it (Gray, 2002). Research has shown that SS are effective in teaching new skills to and decreasing inappropriate behaviors of children and youths with ASD (Croizer & Tincani, 2006; Delano & Snell, 2006; Hagiwara & Myles, 1999; Kuoch & Mirenda, 2003). Teachers and therapists are the primary implementers and only in a few studies researcher-parent partnerships were used (Adams, Gouvossis, Van Lue, & Waldron, 2004; Dodd, Hupp, Jewell, & Krahn, 2008; Kuoch & Mirenda). Either teachers or researchers wrote and implemented SS in these studies. These studies have emphasized that parents, siblings, and grandparents can successfully implement SS written by researchers and positive outcomes obtained in the children’s behaviors. However, to the knowledge of authors there is no study investigating whether family members were able to write their own SS and implement them by themselves.

During the 1960s there was a lack of studies investigating the role of parents as teachers. Teachers/therapists were supposed to deliver intervention in clinical settings to children with ASD in teaching new behaviors and/or controlling their inappropriate behaviors. However, neither maintenance nor generalization effects of these interventions were obtained (Lovaas, Koegel, Simmons, & Long, 1973) during these years. Considering the limited effects of the interventions delivered by teachers in the clinical settings, it was deduced that training family members as teachers could be a model for obtaining positive outcomes regarding maintenance and generalization (Koegel, Schreibman, Britten, Burke, & O’Neill, 1982). Since then, many studies investigating the effects of parent-delivered intervention have revealed that...
parents could deliver intervention successfully and teach their children (Becker-Cottrill, McFarland, & Anderson, 2003; Johnson et al., 2007; Tekin-Iftar, 2008).

When reviewing the studies focusing on parent-delivered intervention it was seen that parents were taught to use effective teaching strategies including discrete trial teaching (Charlop-Christy & Carpenter 2000; Weiskop, Matthews, & Richdale, 2001), pivotal response training (Symon, 2005), and community-based instruction (Tekin-Iftar, 2008) for teaching new skills and appropriate behaviors. The outcomes of aforementioned studies have shown that parents of children with ASD who have limited communication and social skills and repetitive/stereotypic behaviors should especially teach communication and interaction skills and deal with their behavioral problems. Among the studies on parent-delivered SS, no study investigated parent-written and delivered SS in teaching social skills to youth with ASD. In addition to that, the paucity of research investigating effective procedures in teaching youths with ASD is very well accepted. Considering these needs in the literature, we designed a study to test whether family members of youth with ASD could write and deliver SS intervention reliably and youths in the study could learn the social skills that they needed. Maintenance and generalization effects of this intervention were also examined. Furthermore, social validity was examined. Based on the purpose of the study we attempted to answer the following research questions: (a) Do family members learn to write SS for their children correctly and implement it reliably? (b) Is parent-delivered SS effective in teaching social skills to youths with ASD? (c) If youths with ASD learn social skills through parent-delivered SS, will they maintain these skills 1 and 4 weeks after intervention and generalize the acquired social skills across persons and settings? (d) Do family members maintain the acquired SS writing skills over time and generalize the acquired skills on writing new SS for different target behaviors? (e) What were the opinions of family members about the study? (f) Do youths with ASD have similar performance on the social skills taught by their family members as compared to social skill level of their typical peers?

Method

Participants

Three family members (one sister and two mothers) who have youths with ASD in their families and their children (Ali, Umut, Berkant) with ASD were the participants of the study. The first author interviewed the family members about participation in the study. The purpose and procedures of the study were explained and a contract was signed with those who were willing to participate. Then family member-youth dyads were formed. All participants received their ASD diagnosis at a hospital from a pediatric psychiatrist. Gilliam Autistic Disorder Rating Scale-2-Turkish Version (Diken, Ardic, & Diken, 2011) was used to confirm their autism diagnoses. In addition to that Social Skill Rating Scale (SSRS; Gresham & Elliot, 1990) standardized to Turkish by Sucuoglu and Ozokcu (2005) was used for describing their social skill performance. Ali and Umut were below average on SSRS and due to his age Berkant did not take this test. It is applicable to children whose ages are between 3 to 12 years old. Demographics for the family members are presented at Table 1. All participants have deficiencies in social skills.

<table>
<thead>
<tr>
<th>Youths with ASD</th>
<th>Family Members</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Youths</strong></td>
<td><strong>Age</strong></td>
</tr>
<tr>
<td>Ali</td>
<td>12</td>
</tr>
<tr>
<td>Umut</td>
<td>12</td>
</tr>
<tr>
<td>Berkant</td>
<td>16</td>
</tr>
</tbody>
</table>

Settings and Materials

Two different settings were used during the study. Family members’ training sessions were conducted at a meeting room of a Private Special Education Rehabilitation Center where the first author of the study was working. The researchers developed a handbook describing how to write and implement SS with many examples and practices. SS intervention sessions were conducted at each dyad’s own house. Social storybooks written by the family members were used in these sessions. Each sentence is placed on a separate page in the books. Books usually have eight to 10 pages including some visual cues and photos of the youths. Beside these settings, depending on the social skills taught, road from home to school for Ali, school, special education center, and home for Umut and several settings such as shopping malls and streets where target social skill was expected to be performed by Berkant were used as natural settings for testing social skills. A handycam camera was used during sessions throughout the study.

Experimental Design

A multiple probe design with probe conditions across dyads was used in the study. The dependent variable was percentage of correct responses on the performance of the social skill assigned to the youths with ASD. The effectiveness of the SS intervention was demonstrated when the youth was responding at or near to baseline levels during probe sessions and the criterion was reached only after the intervention was conducted.

Target Behaviors

Greeting skill (i.e., saying hello), expressing his happiness or needs appropriately in response to others’ behaviors/comments (i.e., giving a high five, saying thank you, giving a hug, asking to participate in an ongoing activities/game etc), and asking permission for having snacks or various items which his elder sister brings to him (i.e., “May I have the candy bar”) were selected as target skills for Ali, Umut, and Berkant respectively. These social skills were assessed by designing/controlling their environments. For example, Ali’s target behavior was defined as saying “Hello” to the familiar friends/adults coming across during the 10-minute walk to the school. Therefore, coming across five familiar persons on the 10-minute walk to the school was planned for Ali, five occasions which make Umut happy were planned for him, and leaving Berkant’s sister’s handbag open with a candy bar inside alone five times a day in different places for Berkant were planned. Controlled event recording was used during data collection.

General Procedure

Prior to baseline, training of family members was administrated in a group arrangement. Baseline, intervention, maintenance, and generalization sessions were conducted in a one on one instructional format in the study. Prior to intervention, a pilot study was administrated with a sequence of description, modeling, writing, and implementing SS intervention to see whether any modifications/adaptations needed on the program with an 11 year old male with ASD. Except modeling phase, no modification needs appeared during the pilot study.

Parent training. Parent training sessions were conducted in a small group teaching arrangement on a weekend day and took 6 hours. The researcher took pre-post test measures to assess their entry level to the program and acquisition level of writing a SS. The crucial steps for writing SS are presented in Table 2. Table 3 presents the skills needed for testing correct implementation of SS intervention. Presentation of the parent-training handbook took place at the beginning of training sessions immediately followed by asking questions to assess their comprehension level. After ensuring correct responses for each question from each family member, the researcher modeled writing and delivering SS intervention. Modeling steps had two phases (a) modeling for writing a SS and (a) modeling for delivering SS intervention. Family members were required to perform the steps for writing SS presented in Table 2 at 100% accuracy. The quality of the implementation was evaluated by recording the correct answers given by family members. If they did not
meet the criterion on writing a SS, modeling step was re-exposed and continued until they performed the steps of writing a SS with 100% accuracy. After that the process of planning the intervention was initiated. During this process, family members were asked to identify a social skill to be taught. Probe, intervention, maintenance, and generalization sessions for testing the effects of SS were administered by the family members. During these sessions family members were required to deliver five training or probe trials for each social skill and collect data during these trials. They were also expected to perform the steps presented in Tables 2 and 3 for writing a SS and delivering intervention. Writing a SS for a family member took approximately 3 min and implementation of it took approximately 2 min (range: 1 min 59 s and 4 min 29 s). The cost of developing a SS was approximately 2 US dollars.

**Probe sessions.** Full probe sessions were conducted simultaneously with each participant prior to intervention and after meeting criteria in the intervention sessions. A probe session was conducted per day during full probe conditions. Daily probe sessions were conducted to test acquisition after reading the SS. Family members conducted full probe and daily probe sessions in a controlled baseline format. That is to say that the researchers planned/controlled occasions suitable for

### TABLE 2
Crucial Steps for Social Story Writing Skills

<table>
<thead>
<tr>
<th>Steps for Social Story Writing Skills</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Giving a title</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Building the structure with an introduction (description of the situation), a climax (mentioning the reason of the inappropriate behavior, definition of the appropriate behavior and the reinforce), and a conclusion (emphasizing the feelings and thoughts of others against the targeted behavior)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Answering the 5W1H questions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Writing from the point of view of youth</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Using descriptive, guiding, reflecting and/or confirming sentences</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Following the rules for rate of sentences</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Using positive expressions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Writing clear enough for the youth</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Using appropriate sentences and expressions</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### TABLE 3
Steps for Implementing Social Story Intervention

<table>
<thead>
<tr>
<th>Steps for Teaching Skills Using the Social Stories</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Offering the story in appropriate time (Offering the story just before the targeted behavior can be demonstrated)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Offering in appropriate settings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Offering clues (Hi ............... now it is time for reading the story titled ............... which I have written for you. Are you ready?)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Reading the story</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Evaluating whether the story is understood (Asking 4-5 5W1H questions)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Reinforcing the child, if the story is understood</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Ending the process by heading to the milieu (Let's ..................)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Repeating the reading and evaluating process when needed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. If the child is not able to answer the 5W1H questions after reading the story three times, ending the process as a model to the correct answers</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
evoking the target social skills during these sessions. For example, five different persons who are familiar to Ali were asked to be ready on Ali’s school walkway. Five occasions for making Umut happy and satisfied were planned to evoke his response as a consequence. Berkant’s older sister was required to bring a candy bar in her bag and leave this bag open alone in different rooms or places at home. Except for reading a SS during intervention all variables were kept constant. For assessing the participant’s performance level on their target behaviors, Ali was observed by the first author when he goes to school from home. Five persons familiar to Ali were asked to hang around on his way and whether he greeted these persons was recorded by the first author. Probe sessions were conducted at home, school, and the special education center for Umut. Occasions for him to feel happy were created and his responses to these occasions were recorded. Rooms at Berkant’s house were used during probe sessions and whether he asks permission to have candy bar from his elder sister’s handbag was recorded.

**Social story intervention.** SS intervention sessions were conducted just before occasions or opportunities set by the researchers in which target social skills were supposed to be performed. Intervention sessions took place at each participant’s home. Any place including the youth’s room at home where the participants felt comfortable was used as intervention settings. Youth with ASD and their family member sat face to face or next to each other. After delivering attentional cue (e.g., “Berkant, are you ready? Let’s start reading your story.”), the family member read the story and raised comprehension questions (5W1H questions). Correct responses resulted in social reinforcement and incorrect responses resulted in re-reading the story and re-providing the comprehension questions. If a story was read three times and the participant did not answer those questions correctly during a session, then the correct response was delivered verbally and the participant was asked to step into the settings where the target social skills were expected to be performed. Once the participant met the criterion, fading SS reading was initiated. Fading was conducted in the content and the timing of the reading. Fading in the content was implemented in three ways (a) omitting the directive sentence in the story, (b) reading only the title and first-last sentences of the story, and (c) showing the social story book prior to entering the occasions evoking the target social skills. In addition, fading was also planned by increasing the duration between the story reading time and occasions, in which target social skills were anticipated.

**Maintenance.** Maintenance probe sessions were conducted 1 and 4 weeks after the SS intervention. These sessions were conducted just like full probe sessions. Five probe trials were performed. Maintenance sessions were also administered 5 weeks after intervention for testing whether family members maintained writing and implementing SS.

**Generalization.** Generalization across settings and people sessions were conducted by pre-post test measures. Generalization across settings measures was taken for Ali when he was going to his special education center. Generalization data were collected in a shopping mall for Umut and with different persons and settings for Berkant. Pretest generalization sessions were conducted before SS intervention and posttest generalization sessions were conducted after the criterion was met by each participant. These sessions were conducted just like full probe sessions. Generalization was not tested for the participants but for the family members. Family members were asked to write a new SS for teaching different social skills.

**Data Collection and Analysis**

Data were collected to see whether family members would write SS correctly. The steps presented in Table 2 were used to estimate accuracy in writing a SS. The accuracy of writing a SS was estimated by calculating the percentage of correct responses (the number of correct response (step) / the number of total response × 100) presented in Table 2. Inter-observer agreement (IOA) and treatment integrity (TI) data were analyzed. Effectiveness data on acquisition, maintenance, and generalization were collected to examine whether SS written and delivered by family members were effective. Visual and graphical analyses were conducted to examine the effects of the SS intervention. Moreover, maintenance and generalization data for family members were
gathered in the study. Last but not least social validity data were collected to examine whether the aims, procedures, and findings of the study were socially valid. Social validity of the study was tested by social comparison and subjective evaluation. A norm group consisting of 10 males with typical development age 12 to 15 years was formed and their performance on the social skills aimed to teach the participants of this study were assessed. Subjective evaluation was administered to examine family members opinions regarding the study. The researchers developed a social validity question form including 14 closed-ended and four open-ended questions. Social validity data were analyzed by graphical and descriptive analyses.

Results

Findings about Writing SS Correctly

Family members’ skills on writing SS were tested via collecting pre-post test data. Pretest sessions were conducted just before training them how to write SS and posttest sessions were conducted after training to write SS. They were asked to identify a prospective target behavior and write a SS for promoting/changing it. Findings showed that family members were able to write SS with 22% accuracy in the pretest and 100% accuracy in the posttest.

Reliability Estimates

IOA and TI data were collected at least 30% of the sessions which were selected randomly. IOA data showed that at least 97% agreement was obtained across participants and sessions. TI data showed that family members were able to conduct probe, maintenance, and generalization sessions with 100% accuracy. Data also showed that intervention sessions were implemented 100% accuracy by Ali’s and Umut’s mothers and 86% (range = 80%–89%) accuracy by Berkant’s sister.

Effectiveness Findings on SS Intervention: Acquisition, Maintenance, and Generalization

Figure 1 displays data for the effectiveness of family members who delivered SS in teaching social skills to youths with ASD. Data revealed that youth with ASD were able to learn the social skills.

Ali performed his target behavior, saying hello to a familiar person, with 6.7% accuracy during baseline sessions. He met criterion at the third intervention session and showed 100% accuracy across three sessions. Sessions for fading the intervention started at the ninth intervention sessions. Umut performed his target behavior, saying thank you when appropriate, expressing his happiness appropriately, communicating his request kindly, giving a hug to a loved one in a normal magnitude, with 13.3% accuracy during baseline sessions. He met criterion at the seventh intervention session and showed 91% accuracy across the sessions. Sessions for fading the intervention started after having 93% correct performance across the last three intervention sessions. Moreover, he performed his target behavior with 6.7% accuracy during first full probe sessions and after intervention 100% accuracy across subsequent full probe sessions were obtained. Berkant did not perform his target behavior, asking permission to access desired objects/items, during baseline sessions. He met criterion at the eighth intervention and showed 90% accuracy across the sessions. Sessions for fading the intervention started after having 93% correct performance across the last three intervention sessions. He did not perform his target behavior during first two full probe sessions and after intervention 100% accuracy across subsequent full probe sessions were obtained. Fading sessions had three phases and they were terminated when the participants performed at criterion level.

In addition to these findings, all of the participants maintained their target behavior with 100% accuracy 1 and 4 weeks after intervention. Regarding generalization, none of the participants had any correct responses during pretest sessions and after intervention they generalized the acquired target skills with 100% accuracy across persons and settings.

Trend and level analyses were conducted in the study. Immediate effects of the intervention were conducted by subtracting the first data point of the intervention sessions from the last data point of the baseline session. The immediate effects of the intervention were 60
for Ali, 40 for Umut, and 80 for Berkant. These findings showed that social story intervention had positive effects for promoting positive change in the target behaviors. Percentages of nonoverlapping data (PND) were calculated for each participant. For all participants, the PND for the target skills was 100%. These findings showed that family delivered SS intervention had positive effects in promoting change in youths with ASD. Change in

Figure 1. Percentage correct for youths during full, daily, and maintenance probe sessions.
trend direction was evaluated visually in the study and was seen an accelerating change across to adjacent conditions (probe and intervention conditions).

**Maintenance and Generalization Findings for the Family Members**

Maintenance and generalization of SS implementation skill were tested in the study. Maintenance sessions were conducted five weeks after intervention. Data showed that Ali’s and Umut’s mothers maintained SS implementation skill with 100% accuracy and Berkant’s sister maintained the same skill with 86% accuracy. Generalization skills were tested by asking family members to write a new SS for teaching different target behaviors and results showed that family members were able write it with 100% accuracy.

**Social Validity Findings**

Two types of social validity data were collected and analyzed. Youths with typical development age 12 to 15 years were tested on the skills that were planned to teach in this study. Pre-post test data collected from typical youths showed that they performed these skills with 100% accuracy. They were tested on the skill of saying “Hi” to a familiar person on different occasions, saying “Thank you, Oley etc.” in the state of being happy, and giving a hug to someone in a normal magnitude, and asking permission to access some desired objects. Data showed that youth with ASD had significantly lower performance than their peers during pretest on the skills they were supposed to learn. However, after introducing SS intervention their performance seemed to be very similar to their peers. These data are presented in Figure 2.

Subjective evaluation of the opinions of the family members about the aims of the study, the appropriateness of the SS intervention in teaching social skills, and the importance of the findings in their lives were assessed. These data were analyzed descriptively. Overall, family members reported positive opinions on the parameters of the social validity described above.

**Discussion**

Findings showed that not only family members were able to learn writing SS 100% correctly after attending family members training sessions but also they were able to learn to implement SS intervention reliably. Data also showed that SS delivered by them were effective in teaching social skills to youths with ASD, maintaining the acquired skills over time, and generalizing these acquired skills to different people and settings. Regarding social validity of the study, social comparison data revealed that youths with ASD showed similar performance on social skills after SS intervention. Moreover, family members re-

---

Figure 2. Social comparison data for youths with ASD and their typical peers on the social skills before and after social story intervention.
ported their positive opinions regarding the aims, procedures, and the findings of the study. Based on these findings the following points need to be discussed.

To the knowledge of the authors only one study was designed for teaching three teachers of children with ASD to write SS and to implement it on decreasing inappropriate behaviors of their students (Quilty, 2007). Findings showed that although two teachers were able to implement SS intervention with 100% accuracy, the other teacher was able to implement it with only 67% accuracy. A decrease on the inappropriate behaviors was obtained as the effectiveness findings. Besides this study no study was found in teaching parents, siblings, or teachers to write SS to their children/students for the purpose of teaching social skills or controlling their behaviors. There are several studies that were conducted under researcher-parent collaboration. The researchers wrote stories and parents were asked to implement them to their children (Adams et al., 2004; Dodd et al., 2008; Kuoch & Mirenda, 2003). Considering the findings of all these studies, it can be concluded that parents are able to deliver SS intervention reliably and it was effective not only for teaching social skills but also for controlling and/or decreasing inappropriate behaviors. The findings of the present study are consistent with the findings of the previous studies. Therefore this study adds to the literature with its findings. Moreover, when considering the fact that parents wrote SS in this study, a significant contribution is added to the literature.

The following contributions of this study will be discussed by considering the effectiveness findings of the study. As said earlier, one of the most distinguished aspects of this study is about asking family members to write, deliver, and monitor their own SS intervention. In this respect, this study shows that parents can write SS and implement it reliably. This is especially important since it is very well documented that children with autism need intensive and continuous training in their lives. Giving a role to the parents in this respect has valuable contributions in their lives. The other positive aspect of this study is about the ages of the participants of the study. Youths with ASD took part in this study and there is a need in the literature for conducting research with participants who are adolescents. Except for Washburn’s (2006) study with three adolescents with Asperger Syndrome, there is no other study with adolescents.

Social validity assessment was used for testing clinical significance of the study. We used social comparison and subjective evaluation. The findings for both evaluations were highly positive and consistent with each other. Social comparison approach has not been widely used in the literature. Therefore, this study extends the literature on the clinical effectiveness of SS in terms of using social comparison analysis as well as subjective evaluation. Moreover, the findings of subjective evaluation and social comparison supported each other. Another aspect of this study related to the way of fading the SS intervention. The content as well as the timing of the SS intervention were faded in the study. Some sentences in the study were omitted as soon as the participants met the criterion. Also, the duration between reading the story and providing the occasions for the target behavior was increased after the criterion. During the intervention family members read the story just before providing the practice for the social skills and also when criterion was met, fading process took place on the day rather than between the days. The family members read the story at least three hours earlier then when providing the practice. Fading process included three phases in the study and this authentic fading plan can be taken as a contribution to the related literature. One of the points observed during fading process is worthy for discussion. In addition to observed quantitative changes in the performance in the target behaviors of the participants, qualitative changes have also been obtained in the study. During intervention sessions, Ali performed greeting behavior in a feeble voice without using his gestures/mimics; he performed this social skill not only by using verbal greetings but also by using a normal tone of voice and gestures/mimics. Umut used more natural body language and tone of voice when expressing his state of happiness during fading process. The same quality difference was observed in the performance of Berkant as well. He used a normal tone of voice and performed appropriate permission asking behavior during fading process.

Graphical analyses for the experimental sig-
nificance of the intervention were conducted by analyzing level and trend of the data within and between conditions. Within the findings of these analyses, special attention to level analysis is needed. Level change analysis showed that immediate effects of SS intervention are highly positive. This means that as soon as introducing the SS intervention, the level of data across participants has a significant rise towards criterion. The possible reasons for obtaining or at least facilitating these positive changes in the participants can be interpreted by considering the following parameters. First of all, participants do not have intellectual disabilities. Second, the participants might like to use their photographs in the SS book. This could be a motivator for them to read the book and comply with its content. Third, family members were able to identify the target behavior and write SS with high accuracy. Moreover, the most important parameter for explaining high level of behavior change in the participants can be due to the fact that family members were able to deliver SS intervention with a high reliability. On the other hand the characteristics of SS by itself should have a potential power for obtaining behavioral changes. Individuals with ASD perform better in predictable situations, are able to understand and process visual stimuli and prompts and need prompts in order to be socialize with their environment. In light of these characteristics it could be argued that SS allow them to use visual stimuli, include clear messages and define the target behaviors and possible occasions where target behaviors are supposed to be performed.

Data for the maintenance effects of this intervention have positive outcomes. Participants were able to maintain the acquired social skills at criterion level. Maintenance effects of the SS intervention in the previous studies have positive outcomes (Croizer & Tincani, 2006; Kuoch & Mirenda, 2003; Quilty, 2007). The present findings are consistent with the findings of the previous studies.

Two types of generalization data were collected. Generalization was tested on the behavior of family members and youths. Family members were asked to identify a new target behavior for their children and write a SS for obtaining desired change in the target behavior. Generalization for writing new SS for different behavior was tested by pre-post test measures as well. Generalization effects of the SS intervention on the behaviors of the youths were tested in pre-post test measures. In these measures generalization of the acquired social skills to different persons and settings were examined. Findings showed that both family members and the youths were able to generalize their acquired behaviors at 100% accuracy. These findings are supposed to strengthen the external validity of the study. External validity and generalization of the findings of a study can be considered as a significant concern in single subject research designs (Kazdin, 1982). External validity of the findings was obtained through replicating SS intervention within and between youths by introducing and withdrawing SS intervention. Conducting direct replication is another option for strengthening external validity of the findings. Direct replication was conducted by collecting repeated measures within and between youths over time.

Difficulty in generalizing the acquired behaviors is a well-documented problem in training children with ASD. Here, intervention was conducted in the natural settings of participants and in the presence of different people in study. In the past, there was a tendency to teach the behavior and then hope for generalization. But now, it is very well known that systematic planning for promoting/facilitating maintenance and generalization should be developed in the intervention plans. Generalization of the acquired social skills was very high in this study. The possible reason behind this performance can be explained by the settings and persons used in the study. Each social skill was taught in the natural settings where the social skill is supposed to be performed. Also, training sessions were conducted with the presence of different persons that are around the youth with ASD normally. These two points might have positive effects on the generalization performances of participants.

As mentioned earlier, data of this study was evaluated in terms of clinical significance via social validity analyses. Appropriateness of the aims, procedures, and results of the study was investigated through subjective evaluation. Positive outcomes were obtained as a result of these analyses. During this process family
members were asked to share which social skills they would like to teach their children. They preferred to teach the social skills that facilitate their social lives and provide access to the activities and events available around them. The social skills taught in the study were identified based on these opinions. Therefore, it can be hypothesized at the beginning of the study that these social skills may contribute to their daily lives. Hence, social validity of the study in terms of aims and significance of choosing these social skills was established. At the end of the study social validity results showed that family members still reported that teaching these social skills are important in their lives and had positive effects in their quality of life. Furthermore, social validity was tested by social comparison analysis. In order to do that data were collected from same age youths with typical development. As mentioned earlier, initially there was a significant difference between these two groups; however, after introducing SS intervention, their performances looked like each other. When reviewing the previous SS intervention studies in terms of the social validity assessment approach it is seen that 15 studies out of 17 used subjective evaluation (i.e., Adams et al., 2004; Croizer & Tincani, 2006).

In the present study, the performance of the social skills of participants was compared to the performance of the same behaviors of peers with typical development. Ten youths with typical development were tested on pre-post test measures. A dramatic difference was found during pretest measures and after introducing SS intervention posttest measures resulted in same level of performance. Although there seems to be some differences in collecting social comparison data from study to study, the main purpose of these studies is to determine whether participants perform/behave in normal range after intervention. The social comparison findings of the study showed that after the intervention the performance level of the youth with ASD increased to the level of performance of their peers with typical development. Analyzing the social validity of the present study by subjective evaluation and social comparison and obtaining positive outcomes from both findings about the social validity can be regarded as one of the strengths and contributions of this study.

Based on the findings and observations the following recommendations can be suggested. Schools and private clinicians can develop parent education programs similar to one used in this study for enabling them to provide intervention at home to teach social skills. SS intervention can be taught in these programs to parents and/or other caregivers. This strategy, teaching parents as teacher, can help to close the gap between school and home. In addition to that it can help to generalize the skills taught at the schools. The promising outcomes of this study allow us to suggest that the same study can be designed to see the effectiveness of SS intervention in different settings. Also, peer/sibling delivered SS intervention can be planned in the future research. This study is limited with teaching social skills to youths with ASD, and the effects of SS intervention in teaching skills from different developmental areas such as self-care, daily living, and independent living can be planned to examine. The effects of providing SS from mobiles or computers can be investigated. SS were presented in one-on-one arrangement in this study. Future researchers may look at the effectiveness of SS presented via group arrangement.

When reviewing the findings of the previous and present study the following conclusion can be made: (a) Family members were able to learn to write SS correctly and implement it reliably; (b) They were able to maintain these behaviors over time and generalize to write new SS to teach new behaviors; (c) SS delivered by family members are effective in teaching social skills to youth with ASD and also can generalize the acquired social skills across novel situations and maintain it after the intervention; (d) The development and implementation of SS intervention takes a short time and is cost effective; (e) Positive interaction between family members and youths was observed during intervention.

References


Received: 6 January 2015
Initial Acceptance: 12 March 2015
Final Acceptance: 15 June 2015
Abstract: This study evaluated the effectiveness of video modeling with a constant time delay procedure to teach social safety skills to three young women with intellectual disability. A multiple probe design across three social safety skills (responding to strangers who: requested personal information; requested money; and entered the participant’s personal space) and replicated across three participants was used to evaluate the effectiveness of the intervention and participants’ abilities to generalize the skills to in vivo community settings and across novel stimuli. Results indicate that the three participants learned and generalized their ability to verbally respond to perpetrators’ requests for money and personal information, but did not generalize their ability to physically respond to perpetrators entering their personal space.

Teaching personal safety skills to persons with intellectual disability is recognized as an important area of instruction (Dixon, Bergstrom, Smith, & Tarbox, 2010; Kim, 2010; Mechling, 2008). With increased independence and fading of adult supervision in community environments come increased safety risks (Purrazzella & Mechling, 2013) and concerns about these safety risks may interfere with caregivers willingness to support community independence (Purrazzella & Mechling, 2013; Taylor, Hughes, Richard, Hoch, & Rodriguez-Coello, 2004) if proper instruction is not provided. Personal safety skills addressed in the research with students with moderate intellectual disability have included: pedestrian and street crossing safety (Batu, Ergenekon, Erbas, & Akmanaglu, 2004; Brannah, Collins, Schuster, & Kleinert, 1999; Coles, Strickland, Padgett, & Bellmoff, 2007; Collins, Stinson, & Land, 1993; Matson, 1980; Page, Ivata, & Neef, 1976); fire and inclement weather safety (Coles et al.; Mechling, Gast, & Gustafson, 2009; Padgett, Strickland, & Coles, 2006; Self, Scudder, Weheba, & Crumrine, 2007); seeking help when lost in the community (Hoch, Taylor, & Rodriguez, 2009; Purrazzella & Mechling, 2013; Taber, Alberto, Hughes, & Seltzer, 2002; Taber, Alberto, Seltzer, & Hughes, 2003; Taylor et al., 2004); social safety skills - abduction prevention and responding to lures of strangers (Collins, Hall, Rankin, & Branson, 1999; Collins, Schuster, & Nelson, 1992; Gast, Collins, Wolery, & Jones, 1993; Gunby, Carr, & LeBlanc, 2010; Mazzucchelli, 2001); protection against abuse and victimization (Khema, 2000; Khema, Hickson, & Reynolds, 2005); home accident prevention and household safety (Collins & Griffen, 1996; Collins & Stinson, 1994-1995; Feldman & Case, 1999; Jones & Collins, 1997; O’Reilly, Green, & Brauning-McMorrow, 1990; Summers et al., 1992); and first aid skills (Christensen, Lignugaris/Kraft, & Fiechtl, 1996; Gast & Winterling, 1992; Ozkan, 2013). Even though personal safety skills may not be used on a daily basis, they are critical when the need for their use occurs (e. g., escaping from a house fire, not walking in front of a moving vehicle). Although their use is critical, infrequent occurrence is of concern when teaching trials must be limited while waiting on natural opportunities to occur in order to practice the skills (Mechling, 2008). Further, practice of some personal safety skills (e. g., crossing dangerous pedestrian intersections) in a safe environment, prior to in vivo instruc-
tion, may be necessary. It is also recognized that personal safety skills must be generalized across settings and situations where they will be encountered and that programs need to evaluate generalization in natural settings if using simulation for instruction (Mechling, 2008). In a recent review of the literature, Kim (2010) found that few studies used in-vivo assessment to measure generalization of personal safety skills to real life, community-based, situations and therefore, similarly to Dixon et al. (2010) recommended that more in-vivo assessments of generalization be used to evaluate effectiveness of instructional programs that target personal safety skills.

Interestingly, in spite of their recognized importance, few studies exist which have been conducted in the past 10-15 years to teach personal safety skills (Dixon et al., 2010; Kim, 2010; Mechling, 2008) and training of these skills appears to be a neglected area of instruction (Kim). Of particular interest is the application of new technologies that were not available in the late 1900s and early 2000s when the majority of the reported research occurred. One such technology is use of video instruction. Although limited in its evaluation to teach personal safety skills, use of video instruction provides a means to create realistic teaching scenarios in a simulated environment (Branham et al., 1999; Gunby et al., 2010; Mechling et al., 2009; Purrazzella & Mechling, 2013; Tiong, Blampied, & Grice, 1992).

Video technology is a possible way to provide realistic examples of unsafe stimuli (Mechling, 2008) and multiple views of situations that cannot be efficiently created in real life situations (Self et al., 2007). For example, while it may be time consuming to travel to multiple pedestrian intersections for in vivo instruction, video models can provide a variety of scenarios with repetitive teaching trials per scenario. The purpose of the current study was to evaluate use of video modeling to teach social safety skills. Social safety skills for this study were defined as reactionary measures to escape from or end a dangerous situation when it was occurring (Mechling, 2008) and the skills included responding to strangers who: a) requested personal information; b) requested money, and: c) entered the participant’s personal space (e.g., sat too close on a public bus). In addition, a second important research question addressed by the study was whether participants would generalize the skills to naturally occurring scenarios in the community.

The current study further evaluated generalization of social safety skills to real life scenarios without the instructor or any adult, familiar to the participant, being present. In the limited number of studies evaluating generalization of personal safety skills, none were identified in which the researcher or familiar adult was not present. Without this measure, Summers et al. (2011) recognized the uncertainty of whether participants would respond safely without the presence of the investigator and Purrazzella and Mechling (2013) recommended that participants in future studies should travel alone, with no familiar adult present, to evaluate use of a smartphone when lost in the community.

The primary research questions for the current study were: 1) Will video modeling be effective in teaching social safety skills to three young adults with intellectual disability? and; 2) Will participants generalize their behaviors to real life scenarios without the accompaniment of a familiar adult?

Method

Participants

Three females with mild to moderate intellectual disability participated in the study. All were enrolled in a county-funded school system transition program for young adults. The program had three locations and the participants attended the program located on a university campus. Participants were identified by their teacher as students who traveled alone using public transportation and were or would be competitively employed, thus increasing their potential risk for facing socially dangerous situations. Teresa and Lacy had previous experience with video technology and had participated in a prior study evaluating video prompting to complete leisure skills (Ivey, Mechling, & Spencer, 2015). Teresa was 21 years and 9 months old with a diagnosis of moderate intellectual disability and cerebral palsy with left hemi-plegia. Her full scale IQ score on the Wechsler Intella-
gence Scale for Children – Fourth Edition (WISC-IV; Wechsler, 2003) was 45. Measures for adaptive behavior skills could not be located in her permanent record. She read store flyers and simple grocery lists when shopping in the community and used picture and context cues to aid her reading. She cooked with familiar text-based recipes and prepared simple foods without recipes. She recognized buses and their arrival and departure times. She wrote her first and last name in cursive letters and wrote 3-4 word sentences describing her duties while working on spelling words exceeding one syllable. Teresa completed job applications and used a visual model for writing answers to questions such as providing personal references. She used a calculator to add multiple purchases and used the next dollar strategy when paying for items over $10.

Lacy was 21 years and 11 months old with a diagnosis of intellectual disability and ADHD. Her full scale IQ score was 54 on the Wechsler Intelligence Scale for Children – Fourth Edition (WISC-IV; Wechsler, 2003). Her composite score on the Adaptive Behavior Assessment System – Second Edition (Harrison & Oakland, 2000) was 72. She read and followed a daily agenda, read restaurant menus, safety signs, grocery store signs, bus schedules and familiar recipes for meal preparation. She used a computer to locate sale advertisements and read newspaper articles. She wrote grocery lists and was working on nutritional meal planning. She completed job applications with a visual model for more difficult questions and filled out demographic information independently. She wrote and utilized a shopping list to shop and gave cashiers money using the next dollar strategy for amounts over $50. She answered the phone and took messages, but did not own a cell phone. She was encouraged to use a quieter voice when talking and to reduce blaming peers for inappropriate situations as well as increasing telling the truth when describing these situations. Her personal hygiene was improving since becoming employed at Walmart as a store greeter. She enjoyed basketball, using the internet, and listening to music. She traveled alone to places such as the mall using the city bus.

Kimberly was 21 years and 1 month old with a diagnosis of intellectual disability and a seizure disorder. Her full scale IQ score was 59 on the Wechsler Intelligence Scale for Children – Third Edition (WISC-III; Wechsler, 1997). Her composite score on the Vineland Adaptive Behavior Scales (Sparrow, Balla, & Cicchetti, 1984) was 68. She read restaurant menus, recipes, and shopping lists using assistance with pictures. She wrote her first and last name in cursive and was learning to complete demographic information on job applications. She had fine motor delays which affected her writing and had double knee replacements which limited walking long distances. She read a calendar for information and used a personal daily calendar for tracking daily activities. She counted coin combinations with pennies plus nickels and dimes, but was learning to use the next dollar strategy when paying at the cash register. She followed step by step directions from a work list, but needed to increase her initiation of tasks. She volunteered at a day care and thrived when helping with smaller children. She also liked caring for small animals. She displayed social behaviors appropriate for her developmental level and was polite to adults, but had difficulty with peer relationships and frustration. She went into the community in small groups and traveled throughout the community on the city bus with other students. She utilized pedestrian safety skills in familiar settings. She needed to increase her volume when speaking as she often displayed inaudible speech. She enjoyed dancing, vacuuming, and making loom bracelets.
Settings

Video modeling sessions took place in different university classrooms using a portable Lenovo B560 laptop computer. The instructor worked individually with each participant up to three days per week. During video modeling sessions the instructor sat next to the student and advanced the Power Point slides to the next video scenario or to deliver the controlling video prompt. Baseline probe and generalization settings were conducted in various community settings within the university campus or nearby neighborhoods. These sessions took place in locations where participants: a) used the university bus system (on the bus, various bus stops); b) ate at various cafeterias, restaurants, and snack bars and: c) used the library, game rooms, book stores, and recreational facilities.

Social Safety Skill Scenarios and Equipment

Social safety skills were taught using video modeling presented on a laptop. Video captions of a “victim” and a “perpetrator” were made using a Sony Digital HandyCam Camera Recorder and converted into high definition video using Windows Live Movie Maker. Each scenario was saved as a Windows Media Audio/Video file and inserted onto a separate Power Point slide. Three sets of scenarios were targeted for intervention: (a) perpetrator asking for personal information (e.g., “Where do you live?”); (b) perpetrator asking for money (e.g., “Can I borrow a couple of dollars for the shelter?”); and (c) perpetrator entering the victim’s personal space (e.g., sitting too close at a bus stop bench so that the perpetrator’s body is touching the victim’s body). Within each set of scenarios, three different types of question/proximity examples and responses were videotaped and used for instruction [e.g., requests for money for: (a) bus fare; (b) food; and (c) shelter] and within each of the three different types of questions, three different variations of the question were provided (e.g., (a) “I need money for the bus”; (b) “Can you loan me money for the bus?”; and (c) “Do you have money for the bus?” A complete list of the scenarios, question types, variations and responses are listed in Table 1. Three scenarios were used in order to meet the requirements of the multiple probe research design and to target the types of social safety situations frequently encountered in the community. Multiple examples of each scenario and variations of questions/proximity examples and answers were used in an effort to target the multiple forms of questions/proximity examples the participants might encounter in the community and the different responses that would be appropriate. Videotaping and assessments also occurred at multiple locations in the community (e.g., bus stop in front of the Education Building, riding the bus, bus stop at the Hawk’s Nest restaurant). In addition, when assessing for generalization at community sites, one novel type of question/proximity example, with three variations, were tested for each scenario in order to evaluate stimulus generalization (Table 1). These examples were not used during video modeling instruction.

For each video scenario, the model perpetrator approached the model victim (first author) and asked a question or entered her personal space. The model victim then verbally responded or physically moved according to the scenario and example (Table 1). Multiple model perpetrators were used during the videotaping and included both males and females. Video clips ranged from 4s to 20s in duration.

Experimental Design

A multiple probe design across three sets of scenarios and replicated across three participants was used to evaluate participants’ abilities to verbally or physically respond to social safety questions and personal space issues (Gast, Lloyd, & Ledford, 2014). This design was used in order to demonstrate a functional relationship between the intervention (video modeling with a constant time delay procedure) and subsequent changes in students’ verbal and physical responses to social safety scenarios. The study included three conditions: (a) in vivo baseline probes at community sites; (b) video modeling with constant time delay: 0s time delay non-intermixing of different types of question/proximity examples; 3s time delay non-intermixing of different types of question/proximity examples; 3s time delay intermixing of different types of
<table>
<thead>
<tr>
<th>Scenario</th>
<th>Requests for Personal Information</th>
<th>Requests for Money</th>
<th>Entering Personal Space</th>
</tr>
</thead>
<tbody>
<tr>
<td>Set 1:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Video 1:</td>
<td>Live→Response: That is personal</td>
<td>Bus fare→Response: No, Sorry</td>
<td>Sitting too close→Response: Slide over</td>
</tr>
<tr>
<td></td>
<td>Where do you live?</td>
<td>I need money for the bus</td>
<td>Sitting too close on the bus</td>
</tr>
<tr>
<td></td>
<td>Video location: Campus Bus Stop</td>
<td>Video location: On Campus Apartment Building Exit</td>
<td>Video location: Campus Bus</td>
</tr>
<tr>
<td>Video 2:</td>
<td>What is your address?</td>
<td>Can you loan me money for the bus?</td>
<td>Sitting next to at the bus stop</td>
</tr>
<tr>
<td></td>
<td>Video location: Campus Bus</td>
<td>Video location: Outside of Campus Bus Station</td>
<td>Video location: Campus Bus Stop</td>
</tr>
<tr>
<td>Video 3:</td>
<td>Where is your house?</td>
<td>Do you have money for the bus?</td>
<td>Sitting too close on the bench</td>
</tr>
<tr>
<td></td>
<td>Video location: Campus Student Union</td>
<td>Video location: Campus Bus Stop</td>
<td>Video location: Bench outside of Campus Dining Hall</td>
</tr>
<tr>
<td>Set 2:</td>
<td></td>
<td>Food→Response: I don’t have any money</td>
<td>Sit with you→Response: Move to a new table</td>
</tr>
<tr>
<td>Video 1:</td>
<td>Number→Response: I don’t give that information</td>
<td></td>
<td>Joining in at a restaurant table</td>
</tr>
<tr>
<td></td>
<td>What is your number?</td>
<td>Do you have any money, I’m hungry?</td>
<td>Video location: Table at El Cerro Restaurant</td>
</tr>
<tr>
<td></td>
<td>Video location: Outside of Campus Dining Hall</td>
<td>Video location: Inside Einstein’s Bagel on Campus</td>
<td>Joining in at a table in food court</td>
</tr>
<tr>
<td>Video 2:</td>
<td>Can I get your number?</td>
<td>I need a dollar for lunch</td>
<td>Video location: Table Campus Dining Facility</td>
</tr>
<tr>
<td></td>
<td>Video location: Aisle in Lowes Food Grocery Store</td>
<td>Video location: Outside of Campus Dining Hall</td>
<td></td>
</tr>
<tr>
<td>Video 3:</td>
<td>What is your phone number?</td>
<td>Do you have any spare change for food?</td>
<td>Sitting with them at a table in a cafeteria</td>
</tr>
<tr>
<td></td>
<td>Video location: Campus Bookstore</td>
<td>Video location: Campus Dining Hall</td>
<td>Video location: Table Campus Dining Hall Facility</td>
</tr>
<tr>
<td>Set 3:</td>
<td>Name→Response: I’m sorry, I don’t know you</td>
<td>Shelter→Response: Sorry, I can’t help you</td>
<td>Following you→Response: Take 2 steps sideways</td>
</tr>
<tr>
<td>Video 1:</td>
<td>What is your name?</td>
<td>I’m trying to get into the shelter, do you have any money?</td>
<td>Following too closely down an aisle</td>
</tr>
<tr>
<td></td>
<td>Video location: Campus Apartment</td>
<td>Video location: On the Sidewalk outside of Campus Dining Hall</td>
<td>Video location: Aisle in Lowes Food Grocery Store</td>
</tr>
<tr>
<td>Video 2:</td>
<td>You got a name?</td>
<td>I need a place to sleep, can I have some money?</td>
<td>Following too closely down a sidewalk</td>
</tr>
<tr>
<td></td>
<td>Video location: Campus Apartment Halway</td>
<td>Video location: Halway inside School of Education Building</td>
<td>Video location: Sidewalk outside Campus Dining Hall</td>
</tr>
<tr>
<td>Video 3:</td>
<td>What do you go by?</td>
<td>Can I borrow a couple of dollars to get into the shelter?</td>
<td>Following too closely in open area/atrium</td>
</tr>
<tr>
<td></td>
<td>Video location: Campus Apartment Building Stairwell</td>
<td>Video location: Campus Bus Stop</td>
<td>Video location: 2nd Floor Atrium of the School of Education Building</td>
</tr>
<tr>
<td>Novel Questions:</td>
<td>Birthday→Response: Generalize</td>
<td>Baby→Response: Generalize</td>
<td>Stand too close→Response: Generalize</td>
</tr>
<tr>
<td>1</td>
<td>How old are you?</td>
<td>My baby is sick can I have some money for medicine?</td>
<td>Stand too close at a store checkout</td>
</tr>
<tr>
<td>2</td>
<td>Are you 21?</td>
<td>My kids are hungry do you have any spare change?</td>
<td>Stand too close at a bus stop</td>
</tr>
<tr>
<td>3</td>
<td>When is your birthday?</td>
<td>My baby is tired can we have some money to get into the shelter?</td>
<td>Stand too close down an aisle</td>
</tr>
</tbody>
</table>
question/proximity examples; and (c) in vivo generalization probes at community sites. Video modeling was implemented for the first set of social safety skills after the participant’s data showed stable or a decreasing trend across four sessions. Introduction of video modeling was staggered across three social safety skills and a participant’s progression across scenarios was independent of other participants. The order of the social safety skills was alternated across the participants. Criteria for starting a new social safety skill was set at 100% independent correct responding across three sessions when the different types of question/proximity examples were intermixed. Upon mastery of a set of social safety skills, video modeling was withdrawn and four generalization sessions in the community were conducted. Subsequent community generalization probes of mastered social safety skill sets served as maintenance data for the first two sets of skills.

**Dependent Measures and Data Collection**

During baseline probes, video modeling, and generalization probes in the community data were collected on each participant’s ability to verbally respond to perpetrator questions (requests for personal information; requests for money) or physically move away (entering personal space) from the perpetrator. Correct answers and physical movement are described for each set of skills in Table 1. During video modeling an unprompted correct response (anticipation) was defined as verbally answering the question (e.g., “I don’t have any money”) or verbally describing the physical movement (e.g., move to a different table) within 3s of the question or physical movement of the perpetrator presented on the video model. A prompted correct response was defined as verbally answering the question or verbally describing the physical movement after waiting 3s and watching the video model of the correct response (first author saying the correct response to the perpetrator or moving away from the perpetrator). An error or incorrect response was defined as either: (a) an incorrect response before the video model; an incorrect response after the video model; or (b) no response after the video model. Only unprompted correct responses (anticipation) counted towards criteria.

During in vivo baseline and generalization probes, no prompts were provided. Therefore a correct response was defined as verbally answering the question or physically moving away from the perpetrator within 3s of the perpetrator’s question or physical movement into the participant’s personal space. An error or incorrect response was defined as either: (a) verbally making an incorrect statement or physically moving in the wrong direction; or (b) failure to say anything or move in any direction.

**Procedure**

In Vivo Probe 1 (Baseline). Prior to beginning video modeling, each participant was taken to community settings to evaluate her ability to respond to social safety skill situations. During baseline probe sessions participants were brought individually to three different areas of a university campus (Table 1) during each session. At each location a different perpetrator (volunteers secured by the investigators) approached the participant and asked the first question or entered the participant’s personal space from the first example of each of the three social safety skill sets (e.g., “Where do you live?”; “I need money for the bus”; sitting too close on the bus). During the second session of the baseline probe, participants were brought individually to three areas of the university campus different from the first session (Table 1) and were approached by three different perpetrators who asked the second question or entered the participant’s personal space from the second example of each of the three social safety skill sets (e.g., “What’s your number?”; “Do have money, I’m hungry?”; sitting with the participant at the restaurant table). This process continued on the third session whereby questions and personal space were targeted from the third questions/personal space of each social safety skill set. On the fourth session the questions and personal space examples were from the novel examples not depicted in the video models (e.g., “How old are you?”; “Can I have some money to buy medicine?”; Sitting too close to the participant at a checkout line.

At each location, the instructor directed the
participant to the next location (i.e., “Go and wait for the bus and I will catch up with you”; “Go ahead and find a table while I go to the restroom”) and then waited in an area out of the view of the participant and waited until the perpetrator had approached the participant and left. At that time the instructor walked up to the participant, but did not comment on the perpetrator. If the participant made a statement about the perpetrator the instructor re-directed the participant to the activity or provided a vague statement such as “I’m not sure”.

Because the instructor could not hear the participant’s response, the perpetrator recorded the response after leaving the area and also recorded the session on his or her smartphone for reliability measures.

After criteria was reached for responding to one set of social safety skills, probe sessions were again conducted to evaluate the participant’s ability to respond to social safety skill situations not yet introduced through video modeling. These sessions served as generalization measures for mastered sets.

**Video Modeling with Constant Time Delay**

Social safety skills were taught individually using video models of the instructor and different actors who portrayed the perpetrators. The video models were presented using the constant time delay (CTD) procedure and participants were taught to verbally respond to three common types of scenarios. Three sets of scenarios were targeted for intervention: (a) perpetrator asking for personal information (e.g., “What is your address?”); (b) perpetrator asking for money (e.g., “Do you have any spare change for food?”); and (c) perpetrator entering the victim’s personal space (e.g., following too close in a store aisle).

**0s Time delay.** For the first session of each newly introduced set of social safety skills, the instructor turned on the laptop and advanced the Power Point program to the first slide which showed the perpetrator asking a question or entering the instructor’s personal space immediately followed by the instructor’s response (0s delay). Three different examples of the targeted social safety skill were provided with three different responses (Table 1). In addition, three variations of each question/personal space were provided for each of the three different examples for a total of nine trials. The nine trials were then repeated for a total of 18 trials per session. Sessions remained at a 0s delay until the participant responded 100% correct after the video prompt.

**3s Non-intermix time delay.** A 3s delay was then introduced, whereby the instructor paused the video for 3s after the perpetrator asked a question or entered the instructor’s personal space. If the participant did not respond correctly within 3s the instructor played the remainder of the video which modeled the correct response. During these non-intermixed sessions, the three variations of each question type (e.g., requests for money for the bus: “I need money for the bus”; “Can you loan me money for the bus?”; “Do you have money for the bus?”) were presented sequentially. The nine trials were then repeated for a total of 18 trials per session. Sessions remained at 3s non-intermixed until the participant responded 100% correct for one session with the exact response for that type of question (e.g., “No, sorry.”).

**3s Intermix time delay.** The non-intermixed sessions were followed by 3s intermix sessions whereby all three question types and variations (9 total) were presented out of sequence (e.g., “Where is your house?”; “What’s your phone number?”; “What’s your name?”). During these sessions the instructor again paused the video for 3s after the perpetrator asked a question or entered the instructor’s personal space. If the participant did not respond correctly within 3s the instructor played the remainder of the video which modeled the correct response. The nine trials were then repeated for a total of 18 trials per session. Sessions remained at 3s intermixed until the participant responded 100% correct for three sessions using any of the three responses taught for that social safety skill scenario (e.g., “That is personal”; “I don’t give out that information”; “I’m sorry, I don’t know you”).

**In Vivo Generalization Probe**

After a participant met criteria for the video modeling condition with a set of social safety skills, the participant returned to the community settings represented in the video models for three sessions. The fourth sessions of gen-
eralization probes was conducted with three novel questions/personal space examples not depicted in the video models to test for stimulus generalization. The procedures used during the generalization probes sessions were conducted identically to those used during the baseline probe sessions and served as probe sessions for sets not introduced to video modeling.

During the In Vivo Probe 2 condition, sessions were conducted at the sites corresponding to the second variation of each question or personal space example (e.g., Session 1: “What’s your address?”; “Can you loan me money for the bus?”; Sitting too close on the university shuttle; Session 2: “Can I get your number?”; “I need a dollar for lunch”; Sitting with the participant at the food court; and so forth). This pattern of variation of questions continued with In Vivo Probe 3 and during In Vivo Probe 4 the questions and person space examples were identical to those used for In Vivo Probe 1.

Generalization sessions conducted during In Vivo Probe 3 were used to measure maintenance of skills for Set 1 and sessions conducted during In Vivo Probe 4 were used to measure maintenance of skills for Sets 1 and 2 for each participant.

Social Validity
Informal interviews were held individually with the three participants regarding their comfort level in dealing with strangers who might ask them for money etc. and whether they found the video models to be helpful in teaching them these skills. Teresa reported that she wasn’t as afraid to go out by herself on the bus after watching the videos and Lacy said, “No one should mess with me now.” She also said that she liked to watch the investigator in the videos to show her what to do. Kimberly did not appear to understand the question about the strangers, but said, “I think so.” She said that the video models were fun to watch and that she learned from them, “Not to tell strangers anything.”

Interobserver Agreement and Procedural Integrity
During the video modeling condition, interobserver agreement and procedural integrity data were assessed simultaneously by the second investigator on 29.4% of the sessions. Agreement was calculated on a session-by-session basis by comparing both investigators’ data regarding the participant’s response and dividing the number of agreements by the number of agreements plus disagreements and multiplying by 100 (Ayres & Ledford, 2014). Resulting mean inter-observer agreement ranged from 83.3%-100% with a mean agreement of 98.3% across all participants. Mean inter-observer agreement was 100% for Lacy, 99.1% for Teresa, and 96.5% for Kimberly. Investigator behaviors measured for obtaining procedural integrity were: (a) advancing the Power Point slides; (b) pausing the video; and (c) delivering the video prompt after 3s. Procedural integrity was calculated to be an average of 99.7% with a range of 98.1%-100%. Errors occurred when the investigator played an extra slide (extra trial), failure to pause the video, and when the program froze and did not advance to the next slide.

During the in vivo probe sessions, interobserver agreement and procedural integrity data were collected 68% of the time. To measure interobserver agreement, data were collected using a voice recorder application on a Samsung Galaxy S5 smartphone since the investigator was not present when the perpetrator approached the participants. Resulting mean inter-observer agreement was 100% across all participants.

Procedural integrity data were collected by the investigator on whether the perpetrator was in the correct location and asked the correct questions (determined by the video recordings). Procedural integrity was calculated to be 100%.

Results
Figures 1-3 present performance data for each participant across all of the study’s conditions. During the first probe (baseline) condition in the community, Teresa was the only participant who completed correct behaviors in response to the perpetrator. For sessions evaluating the perpetrator entering her personal space she moved over on the bus when the perpetrator sat too close to her (Session 1) and moved to the side when the perpetrator followed her too closely in the aisle (Session 3).
Lacy and Kimberly did not appear to acknowledge the perpetrator entering their personal space and none of the students responded to questions by the perpetrator during the baseline condition. Instead they either stood and stared at the perpetrator or provided him/her with personal information (e.g., “My name is Lacy.”). None of the participants gave the perpetrator money.

When using video modeling with a constant...
time delay procedure, all participants reached criteria in a relatively short amount of time. Kimberly required 6-7 sessions across all three social safety skill sets including Os, 3s without intermixing, and 3s with intermixing procedures. Teresa’s first set of social safety skill scenarios involving a perpetrator asking for money required multiple sessions for her to reach criteria. Even when the questions were not intermixed she required seven sessions to reach criteria across the video modeling instructional sessions. Teresa’s first set of social safety skill scenarios involving a perpetrator asking for money required multiple sessions for her to reach criteria. Even when the questions were not intermixed she required seven...
sessions to reach 100% correct responding. Similarly, when the variations of questions were mixed her performance dropped to 77.8% and she required eight sessions to reach criteria. Her ability to master the personal space and personal information scenarios showed much improvement with only five sessions needed to reach criteria for the personal space scenarios and eight sessions for answering requests for her personal information.

Generalization. All three participants generalized and maintained (where measured) their performance to community settings for the social safety skills of verbally responding to requests for money and personal information. Teresa was the only participant who made an error during the first generalization probe condition following mastery of a set. She failed to answer correctly during the third generalization session when the perpetrator said, “I need a place to sleep, can I have some money?” She did respond by saying, “No”, but this response was not one taught for this set during video modeling. Both Teresa and Lacy maintained their ability to answer questions correctly in the community for sets evaluated and although Kimberly failed to respond to the question, “Where do you live?” during the first session of Probe 4 (maintenance) she
responded correctly to the personal information questions during the last three sessions. All three participants responded correctly (by stating one of the three acceptable responses within a set) to personal information questions and requests for money during novel scenarios not presented in the video models.

Interestingly, none of the participants responded consistently correct when the perpetrator entered their personal space in the community. Kimberly responded correctly during the first session of In Vivo Probe 2 and 3, but failed to maintain her performance during the last probe condition. Teresa responded correctly during two of the four sessions immediately following mastery of the skill, but did not respond correctly during the last three sessions of the last probe (maintenance) condition. The personal space scenarios were the last set taught to Lacy so only one generalization probe condition was implemented. She did not respond correctly to any of the sessions following mastery of the skill with video modeling.

**Discussion**

This study extends the literature on teaching personal safety skills to persons with intellectual disability by demonstrating that video modeling can be effectively used to teach social safety skills in a simulated environment. Social safety skills requiring a verbal response by the participants were generalized to untrained environments and stimuli and maintained over time.

The remaining question is why the participants did not generalize the response to perpetrators entering their personal space in the community. The answer is likely due to a limitation of the study. Although the participants were able to verbally describe what they should physically do (e.g., “Move to another table”) in response to watching the video models physically perform the responses, they never physically engaged in the response during video modeling. It appears from the data that describing these movements was even easier for them to acquire than learning the three verbal responses required when perpetrators asked for money or personal information during video modeling. Most likely this was due to the difference in response requirements between the video modeling and generalization conditions. In order to promote generalization of skills it is important that responses share common physical characteristics similar to those used during teaching (Albin & Horner, 1988). During video modeling, participants verbally responded and during in vivo generalization they were expected to make a gross motor response which did not consistently happen. The results appear to support the findings by Kim (2010) whereby the ability to perform skills in role play situations did not necessarily ensure their performance in applied settings and by Mechling (2008) who reported that a key component, when teaching safety skills was behavioral performance of the skills. Future research using video modeling, in which a physical response is required to social dangers such as responding to the lures of strangers or someone entering the person’s personal space, should consider use of role playing in the simulated environment so that the participants physically rehearse what they see on the video.

The current study adds to the literature by evaluating use of social safety skills in real life situations without the presence of an investigator. Future research may need to evaluate generalization of social safety skills, such as the ones used in the current study, when the participant is traveling completing alone without the investigator being present at any time. Although the investigator in this study was not with the participants when the perpetrator approached them, she did travel with them to different locations along the way during the generalization sessions. This may have provided a sense of security or influenced the participants’ behaviors in some way than if they were totally traveling alone. The design of the study also required that the participants be approached by multiple perpetrators during each in vivo community session. Realistically this would not likely occur in real life situations unless the participant was involved in activities in a particularly unsavory locale.

Despite the limitations, this study highlights an intervention strategy, video modeling, for teaching safety skills that may not convincingly be simulated by other means or practically taught through real life scenarios that may infrequently occur or take place in unsafe situations. As we continue to recognize that
persons with intellectual disability have the right to increased independence and mobility which therefor leads to increased risks, it is important that means for responding to these risks be systematically taught. As stated by Taylor et al. (2010) in response to their own study examining abduction prevention strategies among students with autism, perhaps the current study will contribute toward an increase in attention to effective technologies for teaching safety skills to persons with disabilities.

References


Differentiated Effects of Sensory Activities as Abolishing Operations via Non-contingent Reinforcement on Academic and Aberrant Behavior

G. Richmond Mancil  
Louisiana Tech University

Todd Haydon  
University of Cincinnati

Marty Boman  
Western Kentucky University

Abstract: The purpose of the study was to evaluate the effectiveness of sensory activities used as antecedent interventions on the percentage correct on academic tasks and rate of aberrant behavior in three elementary aged children with Autism Spectrum Disorders (ASD). Study activities were conducted in an after school program for children with ASD where program personnel acted as change agents regarding strategy implementation. An alternating treatment design was used with each participant to evaluate the differentiated effects of three activities. Results varied across participants regarding the sensory related activity that had the greatest effects on producing correct academic responses and reduction in aberrant behavior. In addition, sensory activities had greater effects than control sessions across all participants. A discussion of limitations and future research directions is included.

Autism spectrum disorder (ASD) is a complex developmental disability affecting the lives of over 1.5 million Americans. According to the Centers for Disease Control and Prevention (2014), 1 in 68 children born today will eventually be diagnosed with ASD. In sum, the incidence and prevalence rates of ASD appear to be growing at high rates. Described first by Leo Kanner in 1943 through the case histories of 11 children, these individuals differed significantly from other children; therefore, he recommended that a separate diagnosis was necessary to describe their unique characteristics. Since Kanner’s first description of autism, the disorder has evolved into a spectrum disorder (i.e., ASD) with the percentage of diagnosed individuals increasing each year (Autism Society of America, 2007).

The essential features of ASD include significant impairments in social interaction, communication skills and a highly restricted area of activities and interests (American Psychiatric Association, 1994). The latter category may be more specifically analyzed in terms of restricted, repetitive, and stereotyped patterns of behavior, interests, and activities. Children with ASD often demonstrate a preoccupation with idiosyncratic interests to a level considered abnormal in intensity and focus (American Psychiatric Association, 1994). For example, a child may know about the makes and models of trains and sustain conversations related to this topic for hours, but remain unable to engage in conversations about other topics. Further, many children with ASD have stereotyped and repetitive motor mannerisms (e.g., hand flapping). For example, a child may engage in repeated hand flapping, for no apparent functional purpose such as attention or escape. Several researchers have hypothesized that the function of these stereotyped behaviors are due to difficulties in processing sensory information and may result in other aberrant behaviors which children engage in to regulate environmental stimulation (Baranek, Foster, & Berksen, 1997; Paluszyń, 1979). Significant unusual reactions to various types of sensory experiences in individuals with ASD have been discussed in the literature for decades (Baranek, Wakefield, & David, 2008) and hypothesized to the result of differ-
ences in the brain structure and central nervous system of individuals with ASD. Researchers within the behavior analytic field have purported that behaviors related to these areas are automatically reinforced (Hanley, Iwata, & McCord, 2003; Iwata et al., 1994). When determining the function of these behaviors, social mediation (e.g., access to tangible, attention, escape from task demand) are ruled out (Querim, Iwata, Roscoe, Schlichenmeyer, Ortega, & Hurl, 2013). When this occurs, the function of the behavior is considered to be automatically reinforced either through automatic positive reinforcement (i.e., seeking sensory input) or automatic negative reinforcement (i.e., escaping sensory input).

Individuals with ASD demonstrate social participation challenges that involve functions that are associated with an atypical central nervous system (Bauman, & Kemper, 2003; Courchesne, Carper, & Akshoomoff, 2003). Increasingly, the literature describes the way in which the brain differentiates sensory integration and praxis dysfunction (Crane, Goddard, & Pring, 2009; Dawson & Lewy, 1989; Dawson & Watling, 2000; Smith & Bryson, 1994). Rogers & Ozonoff (1994) report significant incidences of sensory sensitivities and sensory perception deficits in a sample of individuals with autism, suggesting neurological abnormalities in higher cortical sensory perception. Using a meta-analysis of sensory modulation symptoms, Ben-Sasson et al. (2009) reported that 14 different studies have shown sensory differences between individuals with ASD and typically developing individuals with the greatest difference in under-responsivity, followed by over-responsivity and then sensation seeking. Researchers and clinicians have observed changes in persons who seem to react strongly to everyday sensory input, particularly individuals on the autism spectrum who generally have more frequent and intense reactions to external sensory stimuli. Some researchers and clinicians have hypothesized that a person with ASD typically has trouble processing information from the outside world because sensory problems make it difficult to understand what is being seen, heard, and touched. Although the severe reactions to various external sensory stimuli have been discussed in the literature for decades (Baranek, Wakefield, & David, 2008), a systematic process for identifying the precise sensory problems have not been identified and only a few cases of empirical evidence of effective interventions have been (Van Rie & Heflin, 2009).

One study evaluated antecedent exercise effect on behavior maintained by automatic reinforcement (Morrison, Roscoe, & Atwell, 2011). In the study problem behavior decreased during post intervention for three of the four participants; however, the effects could not be attributed to only exercise for one participant. In another study, Saylor and colleagues found that noncontingent auditory stimulation reduced vocal stereotypy in two children with autism (Saylor, Sidener, Reeve, Fetherston, & Progar, 2012). Further, only one study has examined the effects of sensory interventions on academic performance (Van Rie & Heflin, 2009). Van Rie and Heflin employed an alternating treatment design to access the effects of linear swinging, bouncing on a ball, and listening to a story on correct responding on academic tasks. The participants in their study engaged in the activities prior to performing their respective academic tasks. The results for the participants were mixed. For example, one participant’s results were undifferentiated, while others responded better to different sensory interventions. In addition to the mixed results, Van Rie and Heflin noted limitations such as time constraints (break in schedule such as holidays) and changing responses during the intervention. Thus, the purpose of the current study is to extend the research on sensory interventions with individuals with ASD by adding another activity and addressing the noted limitations. In addition, this study is designed to address the sensory issues from a behavior analytic perspective. Specifically, analyzing the access to sensory seeking activities via non-contingent reinforcement procedures to determine if they create an abative effect on off task and aberrant behavior. The current study is designed to determine the effects of specific sensory related interventions, such as linear swinging, and the effect on aberrant behavior and correct response percentages during academic related tasks.
Participants

Selection criteria for participants were based on academic difficulties and sensory challenges. Information on participants was collected by direct observations during instruction/academic work time. Academic samples and information on sensory challenges were collected by indirect interviews from parents and school staff.

Three male participants ranging in age from 8-10 years were recruited from an after-school autism program in central Kentucky that provided services for children and their families. These students were selected because of their difficulty completing academic tasks during instructional periods. They often did not complete and/or did poorly on assignments. Two of the students were in self-contained classrooms and one student was in a general education elementary classroom.

Each child had a diagnosis of Autism obtained independently from a physician or licensed psychologist. In addition, the Social Communication Questionnaire (SCQ; Rutter, Bailey, & Lord, 2003) and the Autism Diagnostic Interview-Revised (ADI-R; LeCouteur, Lord, & Rutter, 2003) was administered to obtain additional scores indicating a diagnosis of ASD. A doctorate level teacher educator and autism specialist trained to conduct the assessments for research purposes administered both instruments to all participants in the study (see Table 1).

Cadmar. Cadmar was a Caucasian male with a chronological age of 8.5 years. Diagnosed at the age of four, he received services from a psychologist (dosage not available), speech therapist (30 minutes a week) and occupational therapist (1 hour a week). Cadmar read at an 8.0 grade level, but struggled with his math skills. He often talked during instructional periods and individual work time. Parents, teachers, and the occupational therapist reported that he had several sensory integration and modulation issues, particularly at school when required to remain in his seat (e.g., bouncing up and down, falling out of his seat). This also was observed in the afterschool autism program. He obsessed about special interests such as dinosaurs and cartoon characters to the extent that he had exhibited their characteristics, such as (e.g., walking around the room in a ‘dinosaur’ motion and ‘growling’ like a dinosaur).

Chuck. Chuck was a Caucasian male with a chronological age of 10.6 years. He was diagnosed with ASD at the age of two, and had been in therapy with a neurologist (no longer receiving), speech pathologist (30 minutes a week) and occupational therapist (1 hour a week) within the past year. His verbal language was limited, but he was able to pronounce a several words to access tangible items such as toys.
or edibles. He was friendly (e.g., hugging everyone, trying to sit in the laps of others), but exhibited challenging behaviors (e.g., spitting) for access to preferred items or to escape task demands. He also had been diagnosed with seizures, which were controlled by medication (the exact medication name was not revealed in school records or from parents). His main sensory issues, according to the parents, occupational therapist, and observed in the afterschool autism program, was spinning in circles around the room.

Hernando. Hernando was a Caucasian male with a chronological age of 9.0 years. He was diagnosed with ASD at the age of three when he did not begin to speak. He had therapy with a speech pathologist (1 hour a week), occupational therapist (1 hour a week). He was quite verbal, but struggled with pragmatics and social situations (e.g., turn taking during board games, initiating conversations). His younger sister also had been diagnosed with ASD, and was non-verbal except for a few utterances. Hernando was very protective of his sister, which resulted in challenging behaviors at times (e.g., hitting other students when perceived to ‘pick’ on his sister). He also enjoyed a number of sporting activities such as soccer and Tae Kwon Do as well as quiet activities including music and movement (e.g., dance to slow music). According to parents and occupational therapists, his main sensory issues included spinning around, hopping up and down while running back and forth in the rooms.

All aberrant behaviors listed were confirmed during direct observations by two independent observers. The observations were compared and reliability of duration and frequency of aberrant behavior between observers was 98%.

Target Behaviors
The target behaviors for the three participants were identified during observations, parent interviews, and analysis of academic permanent products (e.g., math sheets, other data sheets). These included: reading passages for Cadmar, shape and color identification for Chuck, and math equations for Hernando. These activities were of equal high interest for the respective participants, but required further practice for retention. Further, the academic tasks for participants were at the same response effort level across all sessions for respective participants. For example, each reading passage was of similar vocabulary and the same grade level across tasks for Cadmar, while math problems were of similar difficulty across tasks for Hernando.

The target aberrant behaviors differed in topography across participants. Cadmar’s aberrant behavior consisted of walking back and forth in rooms while growling and making ‘dinosaur’ sounds. Chuck’s aberrant behavior involved spinning in circles around the room. Hernando’s aberrant behavior entailed spinning around, hopping up and down while running back and forth in the rooms.

Materials
The materials required for the sensory interventions included a sit and spin, linear swing, and a Hippity hop© ball (i.e., exercise ball with a handle on top).

Independent Variable
The independent variable for this research was the sensory intervention/activity or control activity (i.e., no intervention), which was implemented for five minutes. The three sensory interventions included slow linear swinging, fast bouncing on the Hippity hop© ball,
and slow spinning on the sit and spin. Consistent with the Van Rie and Heflin study (2009), slow linear swinging was conducted with a sling-seat swing attached to the ceiling. The activities chosen were based on ones that may give the same input as the aberrant behaviors for the participants (e.g., spinning in circles—sit and spin, etc.). The research assistant pushed each participant in slow linear patterns for five minutes. For the bouncing ball, participants sat on a 65-centimeter Hippity hop© ball and held the hand on the ball while they bounced up and down for five minutes. Students sat on the sit and spin and used the circular handle in the middle to slowly spin around for five minutes both counterclockwise and clockwise (time for direction was not controlled). Consistent with the Van Rie and Heflin study, a control activity was used to contrast with the sensory interventions. Similar to their study, the participants in this study chose a story-book and listened to a reading by the research assistant (attempting to rule out effects of attention). The control activity was conducted in the same area as the swinging and bouncing activities.

**Dependent Variable**

The dependent variable was the percentage of correct responses on academic tasks (implemented immediately following the sensory intervention or control activity). As previously outlined under target behaviors, each participant completed his/her respective academic tasks. These academic tasks were selected based on participants’ skill development and what each were focusing on in the after school program (determined from review of educational records, parent interviews, and permanent work products).

Additional variables analyzed included off-task behavior and aberrant behavior. Off task behavior was recorded if the participant was not actively working on the assigned academic task for a period greater than 3 seconds. Each aberrant behavior for the respective participants are discussed in the target behavior section.

**Data Collection Procedures**

After the collection of baseline data, each participant was asked to implement the intervention for a five-minute interval session. This was followed by the completion of academic tasks including: a book with comprehension questions, activities with shapes and colors, and math equations.

Reliability checks were conducted by the graduate assistant who was trained to observe and record the occurrences of the target behaviors. Total calculations were completed for each agreement/disagreement for each session between the two observers. Finally, these occurrences were divided by the number of agreements plus disagreements multiplied by 100 (Kennedy, 2005). The mean IOA across all study phases was 95%, 93%, and 94% for Cadmar, Chuck, and Hernando, respectively.

Fidelity for the sessions was recorded by the researcher, who remained consistent for each sensory intervention. The mean treatment fidelity was 100% for Cadmar and Hernando for all three interventions, while Chuck scores were lower at 100% for only two of the interventions (‘sit and spin’ and bouncing ball). During the data collection, he exhibited no change (0%) for the linear swing intervention, he often jumped from the swing and attempted to push the linear swing. Although he was redirected to sit in the swing, he would not remain in the swing for an entire 5 minutes and would engage in challenging behaviors (e.g., spitting, hitting) when redirected to sit in the swing. Thus, the authors determined that forcing him to sit in the swing was not ethically or clinically appropriate. Despite his modifications to one of the interventions, Chuck continued with the academic tasks correctly as outlined in the protocol as reflected by his scores in Figure 3. When Chuck was not provided with any sensory stimulation, he refused to perform any of the academic tasks, which are recorded on the graph.

**Experimental Design**

Prior to conducting the noncontingent reinforcement protocol, a functional analysis was conducted on each individual to determine the validity that the hypothesized automatic reinforcement function was accurate. The conditions included in the functional analysis were: contingent attention, contingent escape, contingent tangible, control (free play), and alone conditions.
Following the identification, an alternating treatment design was implemented to show a functional relationship to the identified intervention compared to the three other interventions (Kennedy, 2005). During the first three sessions, baseline data were collected followed by nine sessions where sensory stimulations was provided and two sessions during which no sensory stimulation was provided according to the research protocol. These final sessions of alternating treatment and withdrawal of stimulus occurred randomly so that the participants were unaware regarding which intervention was going to be implemented each day. All sessions were videotaped and then coded independently by a research assistant and subsequently coded by the first author to check for interobserver agreement (IOA) for 35% or greater of all sessions across all study phases.

**Data Analysis**

This study focused on direct observations of individuals with ASD interacting with various sensory stimuli, and included an alternating treatment design (ATD) with three participants. Baseline data was collected on the participants using behavioral coding of the child’s observed communication, social interaction, challenging behaviors, and academic behaviors during interactions with staff for three sessions. These data were collected during 10 minutes of observation per day, two times per week. For this paper only the academic behaviors are discussed. The staff implemented the various interventions, while graduate assistants collected data during the specified time sequences. During the review of the videotaped session, behaviors were coded using real time collection sheets. The data analysis was completed using Microsoft Excel. Further, permanent products of academic work were collected and checked for accuracy. Data lines were graphed for each intervention, and presented in time-series graphs for each participant. Data analysis was based on visual inspection of the trend of data lines, and magnitude and rate of behavior change between conditions (Kennedy, 2005). Summative data was reported on the fidelity of treatment data.

**Results**

In this study the research evaluated the effectiveness of three sensory activities as antecedent interventions on the percentage correct on aberrant behavior, off task behavior, and academic tasks with three elementary age children diagnosed with ASD.

Prior to implementing the intervention, a functional analysis was conducted on each participant. Results of the functional analysis for each participant were undifferentiated, which indicates the aberrant behaviors were likely
maintained by automatic reinforcement (Smith, Vollmer, & Pipkin, 2007), (see Figures 1, 2, 3). For the intervention, researchers implemented an alternating treatment design to evaluate the differentiated effects of the sensory interventions. Although results varied across participants regarding particular sensory activities that had the greatest effects on the academic correct responses, all participants demonstrated a marked increase in their academic performance as compared to the baseline data after the introduction of the three stimuli (Figures 4, 5, 6). The degree of their performance varied among the participants as well as their preference for one stimulus as compared to another. Further, aberrant behavior for each participant decreased during intervention with the highest decrease corresponding to the same condition as the highest increase in correct responding (Figures 7, 8, 9).

Cadmer. Cadmer completed three sessions for his academic completion baseline until it was determined that the data were stabilized at 40% (Figure 1). This process was followed by three sessions for each of the three sensory interventions randomly administered. Although all three sensory interventions showed an increase in performance, the linear swing was most effective for Cadmer’s academic completion with 100% accurate responses during all three trials of this sensory stimulus. This was followed by an increase with
the sit and spin of 86.7% for these sessions. The Hippity hop© ball intervention demonstrated the least increase of 66.7% (Figure 4) as compared to the baseline although this was a marked increase from the baseline. During these sessions there were no overlapping data points with the baseline data. Further, the most effective intervention as indicated by the data did not overlap with any other intervention or the control. In conclusion, both the linear swing and the sit and spin resulted in academic scores at 80% or higher, which many educators consider acceptable performance.

Chuck. Chuck completed three sessions for his baseline with an average of 10% on his academic completion. The sit and spin intervention sessions were most effective with 100% completion of correct academic responses. Chuck had to be prompted often during the sensory interventions to continue to engage with the task as the protocol required. For example, he chose to walk around the sit and spin with his hands grasping the center turning device rather than sitting on the device with his legs crossed. Despite this modification, he demonstrated an increase in his performance as recorded on Figure 2, which would be classified as “highly effective” (Scruggs & Mastropieri, 1998). During the Hippity hop© ball interventions, Chuck increased to 46.7% from his baseline sessions, and again clas-

![Figure 4. Cadmar’s Percentage of Correct Responses.](image)

![Figure 5. Chuck’s Percentage of Correct Responses.](image)
sified as “highly effective”. Chuck’s performance after the linear swing intervention remained at 16.7% and was considered consistent with the baseline data. This intervention did not show any deviation from the baseline recordings as recorded on Figure 5, and was not effective. Further, the most effective intervention as indicated by the data did not overlap with any other intervention or the control. In conclusion, only the sit and spin resulted in acceptable performance levels, all other sensory interventions resulted in ‘failing’ grades.

Hernando. Hernando’s baseline was established at 26.7% correct responding after three trials as displayed in Figure 3. His results from the implementation of the three sensory interventions demonstrated the most consistency among the data collected across all sensory activities: sit and spin at 90%, linear swing at 80% and Hippity hop© ball at 73.3% (Figure 6). All interventions were in the highly effective range, and indicated that all of the sensory interventions did impact the academic performance. However, only the sit and spin and linear swing resulted in performances considered acceptable for mastery (i.e., 80% or above). Further, the sit and spin was more effective, stable, and ended in an upward trend as compared to the other interventions, but did have one overlapping data point with each of

![Figure 6. Hernando’s Percentage of Correct Responses.](image)

![Figure 7. Cadmar’s Aberrant Behavior.](image)
the other sensory interventions and magnitude in change between the sit and spin and the other two sensory interventions was small.

Discussion

Sensory activities were implemented regularly at the center where this research was conducted. The purpose of this study was to evaluate the effect of three activities (linear swinging, bouncing on a therapy ball and rotating on a ‘sit and spin’) on academic performance (reading comprehension, color and shape identification, solving math equations) and aberrant behavior. Results were somewhat varied as to which sensory stimulation impacted the academic outcomes for the participants. Cadmar had greatest gains in academic performance (100%) when the linear swing was implemented, followed by the sit and spin (86.7%) and the Hippity hop® ball (66.7%). Chuck increased his academic performance on the sit and spin (100%) followed by the Hippity hop® ball (44%). No change was ob-
served with the linear swing as he chose to reject the protocol of sitting on the swing and decided to push it linearly, therefore resulting in no change in academic performance. It should be noted that Chuck did not follow protocol with the Hippity hop© ball as he chose to run around the ball while he held onto the handle rather than sitting on the ball and bouncing. Hernando had greatest gains in academic performance on the sit and spin (100%), followed by the linear swing (80%) and the Hippity hop© ball (70%).

The present study adds to the current research regarding the use of sensory stimulating activities by using non-contingent reinforcement procedures to determine if they create an abative effect on off task and aberrant behavior on the academic performance for individuals diagnosed with ASD. Also, the data from this study give some support to the theory that different stimulation responses results in a variety of outcomes for individuals diagnosed with ASD. Two of the three participants were viewed by the staff of the center to be more excitable than other participants at the center. One participant (Hernando) tended to be hyposensitive to sensory stimulation except when this stimulation was presented in a manner as to impact his sister’s state of mind.

For targeted interventions, this information can be used to explore (a) whether an identified stimulus intervention is likely to succeed in order to increase academic performance and (b) whether minor modifications can be made to assist with other academic tasks. In this study, the results indicated an increase in academic performance; therefore, giving support that these sensory stimuli should be implemented prior to the introduction of new academic information so that individuals identified with ASD can have maximum outcomes.

A few limitations became apparent during this study. The autism program was staffed by individuals highly trained to work with children on the autism spectrum, which could effect overall performance and ease of new interventions being implemented.

Another limitation was the variety of the severity of the diagnosed ASD for the three participants (see Table 1). This was a representation of the heterogeneity of this population, which could have affected the different outcomes for each of the interventions. However, this was addressed somewhat by matching academic tasks to functioning level of respective participants. Also, it should be noted that some of the participants were comfortable with the interventions, while one struggled to follow the protocol for the variety of sensory activities.

Another limitation is the sensory interventions themselves. Each intervention required some level of physical activity; thus, the effects on academic performance could be attributed to engaging in physical activity before an academic task and not the sensory intervention itself. However, since the data indicated a difference between interventions, this may not be the case. Nonetheless, future research should include activities such as running around or some other type of physical activity to rule out these concerns and distinguish between the effects of typical physical activities and sensory interventions.

The current study focused on a targeted intervention that has demonstrated significant advantages for three participants. The simple sensory strategies have positive implications for staff and teachers who work with individuals identified with autism. These strategies can be modified according to the individual’s performance, and can be implemented for a low cost to schools. These results are tentative, but the strategies should be explored in different settings and sensory interventions.

In conclusion, in this study, sensory stimulation suggested that academic performance could be increased dependent on the method implemented; however, the study was conducted with only three participants in one after school program. Further research is needed to determine if the results will be consistent in different settings such as schools. Also, a variety of different sensory stimulations should be presented according to the needs of the individuals to determine the overall impact on academic performance and other behaviors (i.e., challenging behaviors, social interactions).

References


Received: 19 March 2015
Initial Acceptance: 28 May 2015
Final Acceptance: 10 July 2015
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. Three copies of the manuscript along with a transmittal letter should be sent 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.
Visit the official Website of the
Division on Autism and Developmental Disabilities:
http://www.daddcec.org