EFFECTIVENESS AND QUALITY OF EMPLOYMENT SKILL INTERVENTIONS FOR INDIVIDUALS WITH DEVELOPMENTAL DISABILITIES: A META-ANALYSIS, QUALITY REVIEW, AND SINGLE-CASE ANALYSIS

A Dissertation

by

MARGOT BREANNE BOLES

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Jennifer Ganz
Co-Chair of Committee,	Shanna Hagan-Burke
Committee Members	Dalun Zhang
	Victor Willson
Head of Department,	Victor Willson

December 2015

Major Subject: Educational Psychology

ABSTRACT

Many individuals with developmental disabilities (DD) have difficulty transitioning from school to employment due to employment-related skill deficits. There is an increasing need to evaluate the efficacy and quality of interventions that teach skills related to acquiring and maintaining employment. This dissertation includes three articles that assess the efficacy, quality, and individual analysis of single-case studies that implemented employment-related interventions. The first article includes a meta-analysis of 39 studies that met the basic design standards and inclusion criteria for employment-related interventions for individuals with developmental disabilities. Tau-U effect sizes were calculated for each A-B contrast extracted from included experiments. In addition, moderator analyses were also conducted according to the type of intervention, dependent variables, participant characteristics, setting characteristics, and implementer characteristics. Moderate to strong effects were seen across almost all moderator levels and few significant differences were determined between the levels. According to overall effect sizes, video modeling interventions were considered to be effective in teaching employment skills to individuals with developmental disabilities.

The second article includes a review of the quality of 39 single-case studies and 83 individual experiments focused on teaching employment skills to individuals with developmental disabilities. Experiments were assessed and

included in further analyses based on the basic design standards and evidence standards. Study elements were rated according to descriptive design quality indicators. According to the overall analyses, quality video modeling interventions were considered evidence-based according to the 5-3-20 rule.

The purpose of the third article was to investigate the relationship between a point-of-view video modeling (POV) intervention and improvements in socio-communicative skills in two adolescents with ASD. A single-case, multiple-baseline design across three target skills was implemented. Data were recorded for eye contact, body orientation, and withholding interruption using a rating scale. Video modeling clips were recorded with a handheld video camera from the perspective of the participant. Both participants increased appropriate eye contact and body orientation after intervention implementation, but one participant did not show improvement during the withholding interruption intervention phase. Overall, the participants rated POV as useful and satisfactory.

DEDICATION

To my Lord and Savior for giving me perseverance in pursuing the passions you have laid on my heart. Thank you for the peace and joy you have given me through every struggle and uncertainty. My future is uncertain, but wherever I go and whatever I do, I pray that it would glorify you. Philippians 1:6, "Being confident of this, that he who began a good work in you will carry it on to completion until the day of Christ Jesus."

To the families and individuals I have worked with throughout the years. You have opened my eyes and shown me true strength, hope, and perseverance. And to my professors who have supported me through my program. Your patience in guiding me through this process has molded me into the person I am today.

To my family and friends for your constant encouragement and prayers. God has blessed me with a strong, faithful, and loving community.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Ganz for allowing me to learn from you from my senior year as a psychology graduate to fulfilling my doctoral degree in Special Education. I was so blessed to have you as my advisor for all of these years. To my committee co-chair, Dr. Hagan-Burke, thank you for everything you have taught me inside and outside the walls of Harrington. Thank you for giving me the tools to successfully teach and mentor undergraduate students. Thank you Dr. Ganz and Dr. Hagan-Burke for your constant encouragement and patience with me each step of this process. Thank you to my committee members, Dr. Willson and Dr, Zhang, for your guidance and support and thank you for equipping me with the knowledge and skills to be successful in my field.

Thank you to my friends and colleagues, Leslie, Rose, EeRea, and Emily for taking the time to complete IRR for this dissertation. This can often be an exhausting task and I am so thankful for your willingness to help. A special thank you to John for giving me advice and answering all of my questions with patience, often by email at 3am. To my friends and colleagues that I see every day in the halls of Harrington or on my street, Stephanie, Jordan, Derya, Whitney, Jennifer, Heather, Siglia, and Samar. Thank you for your constant encouragement and making life fun even when the stressors of life seemed to be overwhelming.

Thank you to Dr. Grenwelge, Dr. Sumbera, and Christina for your support and allowing me to learn from you, the ins-and-outs of the PATHS career training program that supports successful transitions for individuals with disabilities. This job was a true blessing and a great opportunity.

Thank you to my undergraduate students who I had the privilege to teach. Thank you for showing me how much I love teaching and interacting with intelligent and passionate students.

Thank you to my parents, Becky and Reggie, who have encouraged me and prayed for me continually. You have taught me to love the Lord, love those around me, chase my passions, and work hard in everything I do. I love you with all of my heart. And to my brothers, Mitch, Mac, David, and Dennis, and to my sister-in-law Marcy, for encouraging me and being compassionate towards me when I revealed to you my struggles in completing graduate school. Thank you for taking me to the movies, gathering together for family meals, helping me start a nonprofit, going with me to Aggie games, and letting me cuddle with Boden in order to alleviate dissertation and other graduate school stressors.

Thank you to my extended family for continually praying for me and encouraging me in every step of my program. You have truly been a blessing to me. I love you all.

Finally, thank you to all of my friends here and around the world for your encouraging letters or texts and your constant faith in my completion of this graduate program. I am so proud of you all and how you have applied your gifts

and talents in what you do everyday. You inspire me time and time again to love others and do good in this broken world.

NOMENCLATURE

ASD	Autism Spectrum Disorder
CEC	Council for Exceptional Children
DD	Developmental Disability
DV	Dependent Variable
EBP	Evidence-based Practice
ID	Intellectual Disability
IDEA	Individuals with Disabilities Education Act
IOA	Interobserver Agreement
IQ	Intelligence Quotient
IRR	Interrater Reliability
IV	Independent Variable
NCLB	No Child Left Behind Act
PDD-NOS	Pervasive Developmental Disorder – Not Otherwise Specified
POV	Point-of-View Video Modeling
SCED	Single-case Experimental Design
VM	Video Modeling
WWC	What Works Clearinghouse

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

Many individuals diagnosed with developmental disabilities (DD) face difficulties transitioning from school to employment (Carter, Austin, & Trainor, 2012; Hanley-Maxwell & Izzo, 2012). Fewer than 50% of these individuals are competitively employed after graduating high school, and a majority of those who are employed are unable to retain employment for meaningful periods of time (Newman, Wagner, Cameto, & Knokey, 2009; Newman et al., 2011; Sanford et al., 2011). Many individuals with DD have deficits in a variety of employment-related skills (e.g. interaction with customers and co-workers, task completion, transition between tasks) that make it difficult for them to attain and maintain employment (Grigal & Deschamps, 2012; Hendricks, 2010; Hendricks & Wehman, 2009; Wehmeyer, 1994). Unemployment contributes to (a) dependent living with parents or caregivers, (b) insufficient income to support an independent lifestyle, (c) reduced community involvement, and (d) lower quality of life (Billstedt, Gillberg, & Gillberg, 2005; Hughes, 2001; Wehmeyer, 1994). Unfortunately, many individuals with DD fail to receive adequate transition and employment training services to support their employment-related skills and subsequent employment outcomes (Dolyniuk et al., 2002; Hendricks & Wehman, 2009).

To address the employment-related skill deficits of individuals with DD and provide necessary services, education legislation has expanded its focus to address transition practices and strategies for promoting employment for individuals with

disabilities after graduation (IDEA, 2004; NCLB, 2010). These practices must be implemented no later than an individual's 16th birthday and include: (a) employment driven goals, (b) curricula addressing employment skills, and (c) individualized interventions that support and directly teach employment skills (Alwell & Cobb, 2009; IDEA, 2004).

According to current practices and research, employment skills are typically included in broad definitions of "life skills," "functional living skills" (i.e. skills necessary for daily living), or "transition skills" (i.e. daily living skills specific to the transition period to adulthood; Alwell & Cobb, 2009; Ninci et al., 2015; Palmer, 2014; Wehmeyer & Webb, 2012). The embedding of employment skills within a subsection of a broad skill definition may undermine the importance of teaching and analyzing specific elements of employment skills. Further, the complexity of each employment skill varies according to the type of skill (e.g., social skills, completing all steps of a short or long task analysis), and may impact the individuals' success in acquiring, performing, and completing tasks with fluency (Palmer, 2014). A previous meta-analysis defined employment skills as a combination of relevant social skills (e.g., interacting with customers or other employees, asking others for help, interacting with employers), employment-related task completion (e.g. make copies, assembling an item), and an ability to transition to from one work task to another or to the next step of a task analysis (Wehmeyer et al., 2006). This comprehensive definition of employment skills will be used as a foundation for the meta-analysis and quality analysis in this dissertation.

Employment skill interventions for adolescents and adults with DD are receiving increased attention, particularly among applied research studies with single case experimental designs (SCED) (Ninci et al., 2015; Roth, Gillis, & Reed, 2014; Taylor et al., 2012; Wehmeyer et al., 2006). SCED are unique because of their sole focus on individuals (or small groups of individuals) that serve as their own controls (Gast & Ledford, 2014; Horner et al., 2005; Kazdin, 2011). The primary function of SCEDs is to document a functional relation between an intervention and corresponding target behaviors (i.e., dependent measures) in order to determine the effectiveness of specific intervention components on a participant's behavior over time (Horner et al., 2005). SCED's unique features make it possible to work efficiently with individuals with low-incidence disabilities, and have had a profound impact on the implementation of effective practices in the field of special education (Gast & Ledford, 2014; Horner et al., 2005; Kazdin, 2011).

The growing body of single-case employment intervention research is in need of systematic evaluation to enhance researcher and practitioner understanding of effective practices for individuals with a range of DD (Kratochwill et al., 2013; Ninci et al., 2015; Palmen, Didden, & Lang, 2012; Taylor et al., 2012). Education legislation requires implementation of effective practices that are evidence-based in the classroom, but these prospective practices rarely undergo rigorous statistical and quality analysis (IDEA, 2004; NCLB, 2001; Kratochwill et al., 2013; Palmen et al., 2012). In order to statistically analyze intervention effects, nonoverlap, nonparametric effect size measures are recommended because data from SCEDs rarely meet necessary parametric

assumptions to accurately measure effects (e.g. normal distribution), and because parametric analyses are too sensitive to the presence of outliers in a majority of SCED data (Parker & Vannest, 2009; Parker, Vannest, Davis, & Sauber, 2011b). Previous meta-analyses have used nonoverlap, nonparametric measures including: (a) Percentage of nonoverlapping data (PND), (b) nonoverlapping of all pairs (NAP); and (c) Tau-U (Ninci et al., 2015; Roth et al., 2014; Wehmeyer et al., 2006). The meta-analysis conducted in article one will use Tau-U effect size calculations in order to measure nonoverlap and control for baseline or intervention trend when necessary (Parker, Vannest, Davis, & Sauber, 2011b). To date, no meta-analysis has been published that analyzed employment intervention effects via Tau-U effect size calculations in accordance with moderator variables for individuals with a range of DD.

In addition to assessing intervention effects, there is also a need for assessing the quality of SCED employment intervention studies (Kratochwill et al., 2010, 2013; Taylor et al., 2012). SCED is analyzed differently compared to typical group experimental methods (e.g. randomized control trial) because elements of each study design are typically different (Horner et al., 2005; Kratochwill et al., 2010, 2013). Only a small number of previous meta-analyses and reviews have measured the quality of studies implementing employment interventions (Banda, Dogoe, & Matuszny, 2011; Ninci et al., 2015; Palmen et al., 2012; Roth et al., 2014; Rusch & Dattilo, 2012; Walker, Richter, Uphold, & Test 2010). However, these quality analyses only focus on the basic design and evidence standards provided by *What Works Clearinghouse* ([WWC]; Kratochwill et al., 2010, 2013) or purely descriptive quality indicators (e.g., participant

description, description of procedures, or the simple presence or absence of maintenance phases, generalization phases, and/or social validity measures; Banda et al., 2011; Ninci et al., 2015; Palmen et al., 2012; Roth et al., 2014; Walker et al., 2010). To date, no meta-analysis or review has been published that has combined all suggested quality indicators from multiple sources to produce a more rigorous quality analysis of SCED employment intervention studies.

Social skills (i.e. interviews; interaction with customers, co-workers, or employers; and networking) are considered an important employment skill for attaining a job (Carter & Wehby, 2003; Lee & Carter, 2012; SCANS, 1991; Wehmeyer et al., 2006). In particular, many individuals with ASD have deficits in employment-related social skills including: (a) difficulty giving appropriate eye contact; (b) restrictive interests in conversation topics; (c) difficulty taking turns in a conversation; and (d) difficulty reciprocating in a conversation (Adreon, 2007). Point-of-view video modeling (POV), a form of video modeling, has only been implemented as an intervention with adolescent and adult individuals with ASD to teach independent living skills (Ayres & Langone, 2005; Norman, Collins, & Schuster, 2001; Rayner, Denholm, & Sigafoos, 2009; Shipley-Benamou, Lutzker, & Taubman 2002; Sigafoos et al., 2005). To date, no studies implementing POV to teach employment-related social skills to adolescent and adult individuals with ASD have been published.

In an effort to contribute to the body of research regarding employment interventions for individuals with DD, this dissertation will be composed of three articles that address the previously stated gaps in current research. The first article will consist of

using meta-analytic techniques to measure intervention effectiveness while considering the possible impact of moderator variables. A meta-analysis will allow independent and aggregate measures of effects based on moderator variables for all included articles (Lipsey & Wilson, 2001). The meta-analysis will analyze existing data from independent participants, individual studies, and an aggregate of all studies implementing specific interventions that meet the basic design standards by *WWC* (Kratochwill et al., 2010, 2013). Effect sizes will be calculated using Tau-U in relation to each moderator variable. These moderator analyses were only conducted for video modeling interventions due to the small number of contrasts included in the other interventions. The moderator variables will include: (a) type of target behavior; and (b) task complexity; (c) participant age; (d) participant IQ; (e) participant diagnosis; (f) intervention type; (g) setting; and (h) interventionist.

The second article will consist of a systematic and descriptive quality review of SCED studies implementing employment skill interventions for individuals with DD. Individual elements and overall ratings of the level of quality will be applied to studies that meet or meet with reservations the basic design standards given by Kratochwill et al. (2010, 2013). Other quality indicators will be applied for descriptive purposes only: (a) participant description; (b) setting description; (c) interventionist description; (d) baseline and intervention description; (e) dependent variable description; (f) maintenance; (g) generalization; (h) fidelity; and (i) social validity (CEC, 2014; Horner et al., 2005; Reichow, Volkmar, & Cicchetti, 2008). Visual analysis techniques will also be applied for each data set according to a rating scale that measures evidence of effect

(Kratochwill et al., 2013; Maggin, Briesch, & Chafouleas, 2013). In addition, an analysis of each employment skill intervention, according to the evidence based practice standards (5-3-20), will identify the intervention as evidence based or not (Horner et al., 2005; Kratochwill et al., 2010, 2013).

The third article will consist of a completed manuscript of a SCED study addressing social skills that can be performed in many contexts (e.g. employment, school, home). A POV intervention was implemented for two young men, ages 17 and 19 years old, diagnosed with autism spectrum disorder (ASD). POV was used to teach the participants basic conversational skill that may be used in multiple settings (e.g. school, work, interviews, and community events). This article will discuss the purpose for this study, procedures, results, interpretation of the results, limitations, and implications for research and practice.

Gaps in SCED employment skill intervention literature will be identified in each article to highlight the necessity of the meta-analysis, quality analysis, and SCED study of employment interventions for individuals with developmental disabilities. In addition, each of the following research questions will be addressed in the corresponding articles:

1. What is the magnitude of effect of specific employment skill interventions on acquisition and performance of employment skills for individuals with intellectual and other developmental disabilities?
    a. Is the investigated intervention more effective with (a) certain types or complexities of target skills, (b) certain populations of individuals with

disabilities (i.e., participant characteristics), and (c) certain contexts (i.e. setting and implementer characteristics)?

2. Does the body of SCED on employment skills for adolescent and adult populations with DD meet minimum design and evidence standards as well as adhere to additional quality indicators gleaned from CEC (2014), Kratochwill et al. (2013), Maggin et al. (2013), Reichow et al., (2008), and Wolery (2013)?
3. Is there a functional relation between the implementation of POV alone and improvements in socio-communicative skills for adolescents with ASD?

CHAPTER II

EFFECTIVE INTERVENTIONS IN TEACHING EMPLOYMENT SKILLS TO INDIVIDUALS WITH DEVELOPMENTAL DISABILITIES: A SINGLE-CASE META-ANALYSIS

Many individuals diagnosed with autism spectrum disorder (ASD), intellectual disabilities (ID), and/or other developmental disabilities (DD) have significant deficits in employment skills crucial for job acquisition and maintenance (Grigal & Deschamps, 2012; Hendricks, 2010; Hendricks & Wehman, 2009; Wehmeyer, 1994). These deficits can create obstacles for many individuals with DD when transitioning from school to work, and typically result in unemployment or noncompetitive employment after leaving high school (Carter et al., 2012; Hanley-Maxwell & Izzo, 2012). In addition, these individuals rarely receive the necessary transition and employment skill services to address these deficits (Dolyniuk et al., 2002; Hendricks & Wehman, 2009). Although regulations and resources are in place to support transition into employment, people with DD continue to attain and retain employment at a lower rate than their peers (Wehmeyer et al., 2006; Migliore & Butterworth, 2008). According to successive reports on employment by The National Longitudinal Transition Study-2 (NLTS2, 2015), 30% to 49% of individuals with disabilities such as ASD, ID, and multiple disabilities were currently employed at 4, 6, and 8 years after leaving high school (Newman et al., 2009, 2011; Sanford et al., 2011). This lack of employment for individuals with disabilities can lead to living dependently upon others (i.e., living with parents, guardians, or caretaker),

an inability to earn a stable income, and a decrease in or very little community involvement (Billstedt et al., 2005; Hughes, 2001). These factors lead to self-reports indicating lower quality of life (Wehmeyer, 1994).

In the past 10 years, special education transition practices have developed to include goals, curricula, and interventions focused on teaching employment skills to promote future employment acquisition and maintenance for students with disabilities. The Individuals with Disabilities Education Act (IDEA, 2004) and No Child Left Behind Act (NCLB, 2001) require the use of systematic instruction to teach the skills necessary to prepare students for future employment, beginning no later than when the child turns 16 years old. Employment skills, also referred to as vocational skills (i.e., product assembly, repetitive office tasks, cleaning, participating in an interview), fall into a broader category of "life skills," "functional living skills," or "transition skills" (Alwell & Cobb, 2009; Ninci et al., 2015; Palmer, 2014). While definitions of what is meant by "employment skills" differ, they generally pertain to the accuracy, performance, and fluency of tasks required in the workplace (e.g., completing work tasks correctly and independently, interacting with customers or co-workers, solving problems, and transitioning between work tasks; Palmer, 2014; Secretaries Commission on Achieving Necessary Skills [SCANS], 1991; Wehmeyer et al., 2006). A previous meta-analysis that analyzed specific employment skill interventions suggests that employment-related skills include workplace social skills (i.e., asking for help, solving problems, helping a customer), transition between work tasks, independent or accurate job task completion,

cooking or preparing food as a work task, assembling a product, and cleaning as a work task (Wehmeyer et al., 2006).

Due to the low incidence of many developmental disabilities, single-case experimental design (SCED) has been applied to evaluate promising or existing interventions related to employment skills and special education in general (Gast & Ledford, 2014; Horner et al., 2005; Kratochwill et al., 2013; Ninci et al., 2015; Palmen, et al., 2012; Roth, Gillis, Reed, 2011; Walker, 2010). SCED allows for an analysis of intervention outcomes specific to an individual or behavior due to each individual serving as his or her own control (Horner et al., 2005). Individualized analyses of participants and behaviors provide more information regarding the impact of participant characteristics, target behavior components, overall social significance, and intervention components (Horner et al., 2005). The purpose of SCED is to assess the functional relation between independent (i.e., interventions) and dependent variables (i.e., target behaviors), and as a result, evidence-based practices in this area have been largely influenced by the analysis of intervention effects on target skills via SCED (Gast & Ledford, 2014; Horner et al., 2005; Kazdin, 2011) An analysis of employment skill SCED studies can contribute to the growing knowledge of effective intervention components on socially significant employment skills for individuals with DD.

Independent variables (i.e., interventions) in SCED typically consist of certain behavioral mechanisms and include a variety of procedures and elements based on different theories or strategies (Horner et al., 2005). When an intervention incorporates a unique combination of these elements, this may alter the outcomes. Each of the previous

reviews focusing on functional living skills, including employment skills, analyzed only one intervention category such as, technological interventions (Kagohara et al., 2013; Wehmeyer et al., 2006), community-based interventions (Walker et al., 2010), video prompting (Banda, Dogoe, & Matuszny, 2011), and self-management (Rusch & Dattilo, 2012). Specific interventions, included in the previous employment intervention categories, include (a) video modeling, (b) audio cueing, (c) visual supports, and (d) systematic prompting. These employment skill interventions have only been analyzed individually through SCED and there remains a need to compare effectiveness across multiple interventions. Comparative analyses can assess the efficacy of specific intervention techniques in relation to each other (Lipsey & Wilson, 2001).

Dependent variables in SCED include observable behaviors that are targeted and measured during each condition in a research study (Horner et al., 2005). Previous meta-analyses have focused on broad definitions of functional living skills (i.e., employment skills, independent living skills, social skills, challenging behavior, and self-help skills; Ninci et al., 2015; Roth et al., 2014). Each skill set that falls into the functional living skills definition contains many levels, which makes it difficult to analyze the effects of the intervention related to the complexity or unique behaviors associated with the skill. Employment skills need to be analyzed separately by complexity and types of behaviors involved when analyzing the effects of interventions on behavior change.

SCED is unique because the design is modeled around characteristics and skills of each participant (Gast & Ledford, 2014; Kazdin, 2011; Horner et al., 2005). Participants serve as their own controls during baseline phases and data are then

compared to their performance during intervention phases (Horner et al., 2005). Therefore, it is important to evaluate participant characteristics to determine how this may affect the functional relation between the independent and dependent variables (Council for Exceptional Children [CEC], 2014; Gast & Ledford, 2014; Horner et al, 2005). Participant characteristics that are most commonly considered are diagnosis, age, and functioning level (CEC, 2014; Horner et al., 2005). Previous employment intervention reviews and meta-analyses have solely focused on individuals with a diagnosis of ASD and or individuals with a functioning IQ of 70 or above (Ninci et al., 2015; Palmen et al., 2012; Roth et al., 2014; Taylor et al., 2012). These participant characteristics are very narrow and hinder application of these interventions to a broader population of individuals with disabilities. These reviews call for continued research and analysis of employment interventions for the population of adolescents and adults with a range of intellectual and developmental disabilities in addition to ASD (Ninci et al., 2015; Taylor et al., 2012).

The context of a SCED study includes the location or setting of each condition (e.g., school, home, work), the materials used during the procedures (e.g., smartphone, pictures, toys), and the individuals carrying out the procedures (i.e., interventionist). Previous reviews have categorized the location or setting (i.e., location that the intervention was implemented) as natural (i.e., typical environment of daily life that naturally elicits target behaviors) or simulated (i.e., environments created to mirror the natural environment, but the target skills are not elicited without stimuli similar to those found in the natural environment), but these reviews did not systematically compare

natural versus simulated settings in regard to intervention outcomes (Palmen et al., 2012; Walker et al., 2010). Interventionist characteristics and behaviors (e.g. fidelity when implementing procedures) are additional areas of context that have not been thoroughly investigated by previous reviews or meta-analyses. The setting, the interventionist's level of experience prior to intervention implementation, the role of the interventionist in the participant's natural environment, and the interventionist's implementation fidelity may be important to the effectiveness of the intervention and maintenance; therefore, these elements need to be thoroughly analyzed.

Effect size measures in SCED are crucial to the systematic analysis of the functional relation between independent and dependent variables (Kratochwill et al., 2013). Non-overlap effect size measures are becoming more common when analyzing comparisons of data, and are more accurate for non-parametric analyses of SCED studies compared to parametric measures (e.g., regression; Parker, Vannest, Davis, & Sauber, 2011b). Parametric measures follow certain assumptions (e.g., normal distribution, constant variance) that are not typically present in SCED results, and outliers that are prevalent in SCED heavily impact parametric measures (Parker & Vannest, 2009; Parker et al., 2011b). To avoid errors that would occur when using parametric analyses in SCED, previous meta-analyses have analyzed studies with percentage of non-overlap data (PND), non-overalapping of all pairs (NAP), and Tau-U (Ninci et al., 2015; Roth et al., 2014; Wehmeyer et al., 2006). Tau-U is an effect size that surpasses PND and NAP by measuring non-overlap in compared, adjacent phases while correcting for trend (Parker, Vannest, & Davis, 2011a; Parker et al., 2011b). Tau-U's unique element,

accounting for trend, can create a better definition of the level of intervention effectiveness with more statistical power, unlike regression and other non-overlap methods (Parker et al., 2011b). Though Tau-U is a fairly new measure, multiple meta-analyses have used Tau-U to assess study outcomes (Bowman-Perrott et al., 2013; Camargo, 2012; Ninci et al.,2015; Soares, 2011). Tau-U can provide a more thorough analysis of individual and overall outcomes reported in this meta-analysis.

Many of the existing employment curricula and interventions included in transition plans for individuals with DD have not been put under methodological scrutiny. Employment intervention literature is lacking in overall and comparative analyses for participant characteristics, context, independent variables, dependent variables, and effect size measures (Taylor et al., 2012). A meta-analysis can incorporate these elements for a range of studies with more systematic analyses that strengthen conclusions (Lipsey & Wilson, 2001). This meta-analysis will be the first to use Tau-U effect sizes to measure efficacy of all SCED interventions that meet *WWC* basic design standards (Kratochwill et al., 2010, 2013) and address the teaching of employment skills for individuals with a range of intellectual and developmental disability diagnoses. This meta-analysis considered the following research question in addition to four subcategories:

1. What is the magnitude of effect of employment skill interventions on acquisition and performance of employment skills for individuals with intellectual and other developmental disabilities?

a. Is the investigated intervention more effective with (a) certain types of interventions, (b) certain types or complexities of target skills, (c) certain populations of individuals with disabilities (i.e., participant characteristics), and (d) certain contexts (i.e., setting and implementer characteristics)?

Methods

Search Procedures

Articles for possible inclusion were identified using electronic searches of the following databases simultaneously: *(a) Academic Search Complete, (b) Applied Technology Full Text, (c) ERIC, (d) Education Full Text, (e) Professional Development Collection*, *(f) Psychology and Behavioral Sciences Collection, (g) Social Science Full Text, (h) Vocational and Career Collection,* and *(i) Vocational Studies Complete*. There were no specifications for date of publication. Key search terms were separated into two groups and then combined with each other. Each search term in group one was paired with every word from group two by the word *and*, creating a Boolean search. The first group contained the words *autis*, Asperger*, ASD, PDD*, pervasive developmental disorder, development* disab*, low-incidence dis*, intellectual* disab*, mental* retard*,* and *multiple disab**. The second group contained the key words *employ*, career*, vocation*, employ* skill*, career skill**, *vocation* skill*,* and *job skill**. An asterisk was added to many search terms for a more encompassing search due to the variations of these words when referred to in published articles (see Table A1 in Appendix A for the total number of sources gleaned from each search combination).

Inclusion and Exclusion Criteria

Title and abstract inclusion/exclusion. To create a clear objective for the review, a set of inclusion/exclusion criteria were applied to the title and abstract of each article. These criteria consisted of (a) SCED, (b) at least one participant diagnosed with a DD, (c) the dependent variable measures consist of transition skills (i.e., independent living skills, employment/vocational skills, social skills), (d) an intervention component is the independent variable, and (e) a journal article or dissertation published in English. Sources with SCED was an important criterion because SCED focuses on the individual and intricate details in study procedures and interventions. SCED has informed special education in individualized, promising, and evidence-based practices throughout the years, particularly given low incidence populations (Gast & Ledford, 2014; Horner et al., 2005). The disability diagnosis is important when focusing on populations of individuals that may result in intervention outcomes due to population characteristics. Sources containing an intervention as the independent variable was needed as a criterion because the entire meta-analysis is based on the functional relation between the independent and dependent variables. In order to analyze the functional relation, an intervention had to be implemented systematically in the study. The criterion for the target dependent variables (i.e., transition skills) was broad in this abstract/title screening, even though the focus of this meta-analysis was employment skills, in order to avoid premature exclusion of a source if the target skill was questionably linked to employment skills. Finally, even though employment skill interventions were more geared towards adolescent or adults with disabilities, an age criterion was not applied

during the title and abstract inclusion/exclusion screening because individuals under the age of 12 could have been involved in pre-employment skills training. The inclusion/exclusion criteria were broad in this phase due to the limited information given in titles or abstracts of the articles and to avoid the exclusion of articles that could have been included with further analysis. If the title and abstract did not include enough information to be confidently included or excluded based on the criteria given above, the entire article was downloaded and the inclusion and exclusion criteria were applied to the full text.

Full text inclusion/exclusion. After excluding articles based on the information extracted from the titles and abstracts, full text criteria were applied to all of the remaining articles. The criteria consisted of (a) a SCED (i.e., reversal/withdrawal, alternating treatments, multiple baseline, multiple probe, multi element design); (b) at least one participant diagnosed with a DD; (c) employment/vocational, or pre-employment/prevocational skills measured as dependent variables; (d) an intervention implemented to improve and promote independent performance of employment skills; and a (e) line-graph representation of time-series data regarding task completion and independent performance of target skills (e.g. number of prompts needed, percent of steps completed independently). Employment skill dependent variables and intervention inclusion specifications were applied to focus this review on interventions targeting employment skills. The last criterion, line-graph representation of data, was applied to aid in data extraction visual analysis.

Design Standards. The last round of inclusion and exclusion criteria required the analysis of each experiment (i.e., each single-case example) included in a study using the basic design standards according to *WWC* (Kratochwill et al., 2010, 2013). Studies that met or met with reservations the basic design standards were included in this meta-analysis. The design standards, based on Kratochwill and colleagues' (2013), included: (a) the purposeful manipulation of the independent variable, (b) recording the overall interrater reliability (IRR) or interobserver agreement (IOA) for 20% of sessions and with an overall score of at least 80% agreement (each element of IRR/IOA was scored separately), (c) three different attempts to present an effect for the independent measure occurred at three separate points in time, and (d) 3 to 5 data points were present in each relevant phase.

Each of these basic design standards fulfills a specific role in assessing the quality. Systematic manipulation requires the researcher or interventionist to change components of the conditions in which the intervention is implemented in an organized way (Kratochwill et al., 2013). This allows for researchers to control for confounds in the environment that may skew the results. IRR/IOA provides dependent variable measures agreement between two raters or observers across time, which solidifies the reliability of data collection methods and results (Horner et al., 2005; Kratochwill et al., 2013)**.** It is very important for researchers to attempt to present an effect multiple times and at multiple points in time to strengthen experimental control (Horner et al., 2005; Kratochwill et al., 2013). Attempts to present effect need to occur at three different points in time to promote replicability of the effect and a more concrete conclusion that

the intervention, not the environment, is attributed to the change (Kratochwill et al., 2013). The number of data points included in each phase is important to presenting effect and representing consistency in the data (Kratochwill,et al., 2013). Five data points is ideal when analyzing the quality of the study because they provide more information regarding the trend and consistency of the data (Kratochwill et al., 2013).

When reviewing the potential articles for inclusion, each design standard was scored and then an overall score was given for each study according to a modification of the rating system used in Maggin et al. (2013), influenced by *WWC* (Kratochwill et al., 2013). The overall scores included meeting the design standard (score of 2), meeting the design standards with reservations (score of 1), or not meeting the design standard (score of 0). Boles (2015a, available online) provides the ratings and descriptions of each standard. Please note that these design standards apply separately to each experiment (i.e., all single-case designs represented in each study; Kratochwill, 2013), therefore, when the design standards did not meet the minimum standards, that experiment's data was excluded while another participant from the same study (if he or she possess a data set that meets the minimum standards) was included in moderator analyses.

Ancestral search. Following the design standards screening, an ancestral search was implemented with the previously included articles. An ancestral search included searching through all of the references given in the included studies and running these references through inclusion and exclusion processes. Forty-three studies for possible inclusion were reviewed via the full text inclusion/exclusion criteria. The remaining studies were then analyzed for inclusion based upon the basic design criteria

(Kratochwill et al., 2013). With the completion of all screening processes, the remaining studies were included in this meta-analysis. A summarized list of the number of articles remaining after every screening process is reported in Appendix A (see Figure A1).

IRR for inclusion/exclusion. Two raters independently screened 30% of the sources accumulated after the database search for the title and abstract and full text inclusion/exclusion process. The sources were randomly selected from the total accumulation of sources to avoid selection bias. The basic design standards inclusion/exclusion screening was completed by two raters for 50% of the articles. When a disagreement occurred in any of the inclusion and exclusion processes, two raters discussed the discrepancy and reached a consensus. IRR for inclusion/exclusion screenings were calculated via simple percent agreement (total number of agreements divided by agreements plus disagreements and then multiplied by 100) and Cohen's kappa (Cohen, 1960).

Coding

Descriptive coding and moderator coding based on the characteristics of each study included (a) intervention type, (b) type of target behavior, (c) task complexity, (d) participant age, (e) participant IQ, (f) participant diagnosis, (g) setting, and (h) interventionist (see Appendix B for moderator codes). In the case that a participant, intervention, or other characteristic did not fit into the listed levels (e.g., age not provided), the study characteristic was labeled as OTH (i.e., other) and excluded from that specific moderator analysis.

Intervention type (independent variable). The type of intervention is very important when considering the most effective and efficient tool to aid an individual in reaching desired and necessary employment goals. Common interventions for teaching functional living skills, including employment skills, have included video modeling, audio cueing or coaching, visual schedules or scripts, and prompting procedures (Ninci et al., 2015). Video modeling interventions included a video or live modeling of a target skill by an individual (e.g., the individuals, peer, adult). After the participant watched the video or live modeling, the target skill was performed. Audio cueing/coaching included pre-recorded or live prompts via the interventionist from a distance while the participant performed the target skill. These audio cues are typically delivered via earpieces or technological device. Visual interventions included static pictures/words or scripts individually or in a group of like items that represent the target skill as a whole or each step of the target skill. Prompting interventions included systematic use of more or less intrusive prompts to promote independent performance of the target skill. Knowing which interventions are the most effective or which ones are promising strategies can give teachers, practitioners, and researchers a more focused approach when preparing individuals with disabilities for the workforce. This moderator included 4 levels: VM (video modeling which includes: video prompting, video priming, video self-modeling, point of view video modeling, adult/peer video modeling, or in-vivo modeling), AC (audio cueing or audio coaching), VIS (any visuals, static pictures, and picture or written schedules or scripts), and PRMTS (most-to-least or least-to-most prompting implemented alone or as the primary intervention). Based on the results of the

intervention moderator analysis and the larger number of studies and contrasts analyzed, studies that implemented video modeling interventions were analyzed according to the moderators described below.

Dependent variables. Each video modeling study was analyzed according to the type of skill targeted by the intervention and the complexity of the skill.

Target behavior. The type of target behavior is important due to the range of skills included in the definition of employment skills. This moderator aids in answering the question: are employment interventions just as effective or more effective when teaching different types of target skills? This moderator informs teachers, practitioners, and researchers what skills are best taught with certain interventions. The target behavior moderator separated job specific skills from employability skills. Job-specific skills are only relevant to the current place of employment and this skill would be difficult to generalize to another employment setting. Employability skills can generalize across employment contexts because the skill is taught broadly and can be used when performing a variety of job tasks. The target behavior moderator included five levels: SS (i.e., social skills such as requests, interview skills, interacting with customers, asking questions; employability skill), CLEAN (i.e., cleaning the workplace, cleaning a certain product; job-specific skill), ASSEM (i.e., assembly or packaging, stuffing envelopes or folders, putting together more than one piece of an object; job-specific skill), REP (i.e., repetitive tasks that do not require assembly and occur more than once in one sitting such as, copying, faxing, or stocking; job-specific skill), and TRANS (i.e., transition from tasks or task steps; employability skill).

Task complexity. Each employment skill has a specific level of complexity according to the amount of steps in the task analysis (e.g., making a copy) or difficulty involved in simply performing the skills (e.g., social skills). This possible moderator gives teachers, practitioners, and researchers insight into the effectiveness of interventions depending on the complexity of the employment skill being taught. The task complexity moderator included five levels: LOW (i.e., low complexity of 1 to 10 steps involved in the task analysis for a specific task), MOD (i.e., moderate complexity of 11 to 20 steps involved in the task analysis for a specific task), HIGH (i.e., high complexity of 21 or more steps involved in the task analysis for a specific task), SS (i.e., social skill tasks, high complexity), and TRANS (i.e., transition from task to task). Tasks that did not match the previous codes or were not specified were placed in the OTH code category. All interventions in the OTH category were not included in the analyses.

Participant characteristics. The participant age, diagnosis, and IQ variables were coded for video modeling studies to give insight into the effects of participant characteristics on the functional relation between various interventions and employment target skills.

Age. The participant age consisted of three variable levels: MS/HS (i.e., middle school and high school students ages 12 to 15), HS (i.e., high school students ages 16 to 21), and ADULT (i.e., ages 22 and over). These levels were created based on the ages of individuals in specific stages of school curriculum and transition planning. IDEA (2004) requires transition and employment skill planning to begin no later than the individual's 16th birthday. Due to this requirement, the variable levels include an age group (i.e.,

MS/HS) that does not require transition planning, but may be granted transition planning by an IEP team; an age group of students (i.e., HS) at the beginning of transition planning through high school; and an age group of adults (i.e., ADULT) who have graduated high school and are not receiving transition services anymore. This age variable gave insight into the effectiveness of transition planning according to the age limits specified by IDEA (2004).

Diagnosis. Participant diagnosis, gives insight into the effectiveness of employment skill interventions on target skills for specific populations. The interventions can be more effective for individuals diagnosed with ASD than those with ID, or there may be no difference at all. There were three levels for this variable: ASD (i.e., autism, autism spectrum disorder, Asperger's, PDD-NOS), ASDID (i.e., autism spectrum disorder with an intellectual disability), and ID (i.e., intellectual disability).

IQ. The last participant characteristic is IQ. This variable was broken up into two levels: HIGH (i.e., score of 50 or above) or LOW (i.e., score of 49 or below). These levels were chosen due to the definitions of the severity of the intellectual disability according to the IQ score consisting of mild intellectual disabilities, defined as scores at 50 or above; and moderate to severe intellectual disabilities, defined as scores below 50. The definition of intellectual disability also includes deficits in adaptive behavior (i.e., daily living skills), but due to the lack of adaptive skill assessments included in the relevant studies' participant descriptions, this variable only included reported IQ measures. Intellectual functioning and adaptive skill deficits can play a role in the effectiveness of employment skill interventions.

Setting. The moderating effect that the setting may have on each video modeling intervention's effectiveness can provide necessary information regarding the implementation of an intervention in a simulated setting versus a natural setting. Many school and employment programs used simulated training settings that contained materials and stimuli included in the employment setting due to a lack of resources (e.g., time, funds, and/or employment locations) and/or to promote efficiency (Lattimore, Parsons, & Reid, 2006; Test, Richter, & Walker, 2012). In contrast, employment skill training is often preferred in a natural setting (i.e., job-site, on-the-job, or community-based training) with materials and stimuli that are naturally present in that environment to promote generalization (Lattimore et al., 2006; Test et al., 2012; Walker et al., 2012). Setting was coded in two levels: SIM (i.e., a simulated setting is not in the typical setting where the job task is performed, but does include materials and stimuli similar or identical to those found in the employment setting; e.g., a classroom with items to assemble on each desk, a school office with a fax machine, or a training room with office materials), and NAT (i.e., a natural or employment setting where the skills being taught would typically occur with naturally occurring materials and stimuli; e.g., office building, department store, warehouse).

Interventionist. The interventionist is an important moderator variable due to the fact that some interventionists might be more experienced in implementing video modeling interventions (i.e., more precise treatment fidelity); therefore, making the interventions more effective. This moderator is also important because it is beneficial to know if other interventionists with less training, who are present in the participant's

natural environment (i.e., school, community, or place of employment), can implement the intervention correctly and see benefits in target behaviors while promoting maintenance and generalization of the skills in the natural environment. There were three levels for the interventionist variable: TCHR (teacher or instructor), JC (job coach), and RSCHR (researcher or experimenter). Interventionist descriptions that do not match the previous codes or are not specified were placed in the OTH code category. All study descriptions in the OTH category were not included in the moderator analysis.

Interrater reliability for coding moderators. All (100%) of the included studies were independently coded by two raters for the aforementioned moderator descriptions. In the instance of a disagreement two raters discussed and came to a consensus. IRR was measured according to simple percent agreement.

Raw Data Extraction

Raw data were extracted from each included article's line-graphs that contained an intervention targeting employment skills for individuals with a developmental disability and met or met with reservations the basic design standards. Data extraction focused solely on the baseline and intervention phases that were adjacent to each other. Each included data point was rank ordered as suggested by Parker and colleagues (2011a), starting with 1 as the lowest data point and going up by one point for the next level including one or more data points. If the target behavior was hypothesized to decrease (i.e., job coach prompting), then the rank order was determined by starting at the highest data point (rank of 1) and increasing as the levels moved downward. For example, in a multiple baseline design including at least three different behaviors and an

intervention that is introduced to these behaviors at different points in time ($A_1B_1A_2B_2A_3B_3$), each A and B phase pair would be rank ordered separately. The data extraction process included taking a screen shot of the included graph, pasting this screen shot in a spreadsheet file, electronically drawing straight lines that match the level of each included data point, analyzing the rank order of each included data point, and recording the list of rank orders for each included phase.

Calculating effect sizes. Tau-U is an effect size based on the non-overlap of data in phase contrasts while accounting for trend in baseline, if needed (Parker et al., 2011b). Tau-U measures range from -1.0 (indicating a reverse of hypothesized effects) to 1.0 (indicating no overlap between phases and 100% improvement in the intervention phase compared to baseline; Parker et al., 2011b). Tau-U was calculated for each phase contrast according to each indicated moderator via a Tau-U calculator (Davis & Davis, 2014). In addition to calculating Tau-U and to create a more conservative analysis of the moderator variables, the Kruskal-Wallis one-way analysis of variance was conducted to measure statistical significance (Kruskal & Wallis, 1952) or Wilcoxon statistics (Wilcoxon, Rhodes, & Bradley, 1963) for comparison of variables with only two levels. If the Kruskal-Wallis measure showed statistical significance, the Dunn post-hoc test was implemented to analyze existing pair-wise comparisons (Dunn, 1964). The only exceptions were the IQ and setting moderators with only two levels per moderator; therefore, it was necessary to run the Wilcoxon two-sample test for significance instead of the Kruskal-Wallis (Wilcoxon et al., 1963). Study elements that were coded as OTH were excluded from analyses due to the heterogeneity of those levels.

Interrater reliability for data extraction. All (100%) included graphs were independently rank ordered and visually analyzed by two raters. IRR was attained for all included graphs due to the high possibility of error from inattention to minute details or difficulty analyzing the graphs because of distortion due to poor publication resolution. IRR was measured according to simple percent agreement. In the case of any disagreements in the data extraction process, two raters discussed the discrepancy and reached a consensus.

Results

This meta-analytic review analyzed 235 AB contrasts from 39 studies with a total of 120 participants. Descriptions of strong (i.e., 0.85 – 0.95), moderate (i.e., 0.70 – 0.84), or weak effects (i.e., below 0.70) were used to categorize Tau-U effect sizes for all items discussed below (Ninci et al., 2015). Tau-U scores ranged from 0.34 to 1.00 for each study and -0.67 to 1.00 for each contrast. A total of 29 studies and over 180 contrasts resulted in independent Tau-U scores of 0.85 or above, resulting in strong intervention effects on acquisition and performance of employment skills. Due to the wide range of study components (e.g., participant characteristics, intervention type, and target behaviors), a Tau-U moderator analysis and significance testing ($p < 0.05$) were conducted for each intervention type to determine which studies would be included in the remaining moderator analyses (see Appendix B for moderator codes). Tables C1 to C5 in Appendix C provide Tau-U scores and significance scores (if applicable) for each moderator level given below.

Intervention Type

Four types of primary interventions that were most commonly implemented in a majority of studies (i.e., video modeling, audio cueing, visual, and prompting) were analyzed individually and comparatively. Tau-U scores for the type of intervention implemented ranged from moderate effects of 0.83 CI_{95} [0.79, 0.87] for video modeling interventions to strong effects of 0.93 CI_{95} [0.84, 1.00] for prompting interventions, 0.97 CI_{95} [0.91, 1.00] for visual interventions, and 0.97 CI_{95} [0.85, 1.00] for audio cueing/coaching interventions. The Kruskal-Wallis test did indicate a significant difference between moderator levels ($p = 0.0001$). The Dunn post-hoc test indicated a pairwise significant difference between the efficacy of video modeling and audio cueing/coaching interventions. Audio cueing/coaching interventions were shown to have stronger effects than video modeling interventions, but audio cueing interventions only included 4 studies and a total of 17 A-B contrasts; therefore, these results should be interpreted with caution. There were no other significant differences between the other interventions. All Tau-U scores and Dunn post-hoc results pertinent to the intervention moderator analysis are reported in Table C1.

Studies implementing video modeling interventions yielded the highest number of contrasts ($n = 112$); more than double the contrasts yielded by studies implementing visual interventions ($n = 51$). Audio cueing/coaching, visual, and prompting interventions yielded small numbers of contrasts, making it difficult to break down overall effect sizes by additional moderators. Studies implementing video modeling interventions not only contained enough contrasts to run additional moderator analyses,

but the overall moderate Tau-U scores calculated for this intervention warranted further analysis. As a result, moderator analyses for types of target behaviors, complexity of employment tasks, characteristics of participants, types of settings, and characteristics of interventionists were conducted with studies that implemented video modeling interventions. Significant differences between the moderator levels were conducted with Kruskal-Wallis, Wilcoxon two-sample for moderators with only two levels, and Dunn post-hoc tests for moderators with three or more levels.

Dependent Variable

Target behaviors. Tau-U scores for specific target behaviors ranged from weak effects for job-specific repetitive target behaviors to strong effects for job-specific assembly and cleaning behaviors as well as both employability skills (i.e., social skills and transition skills; see Table C2). Job-specific cleaning behaviors yielded the strongest effect 0.92 [0.76, 1.00], but results should be interpreted with caution due to the small number (n = 5) of contrasts. Job-specific assembly skills (0.89 CI_{95} [0.80, 0.98]), and employability social skills (0.85 CI_{95} [0.78, 0.92]) yielded strong effects from 21 to 31 contrasts respectively. Overall, assembly and social behaviors yielded strong effects with narrow confidence intervals and a larger number of contrasts per level; therefore, these results can be interpreted with more confidence compared to the small number (n = 5, 16) of contrasts for cleaning and transition behaviors.

The Kruskal-Wallis did indicate a significant difference between target behavior moderator levels (p = 0.0023). The Dunn post-hoc yielded a pairwise significant difference between the efficacy of video modeling interventions for assembly versus

repetitive target behaviors. Assembly behaviors were shown to have stronger effects than repetitive behaviors (see Table C2).

Task complexity. Tau-U scores for task complexity ranged from moderate effects 0.71 CI_{95} [0.62, 0.80] for moderately complex target behaviors to strong effects for 0.88 CI_{95} [0.75, 1.00] for lower complexity transition skills (see Table C3). Complex social skills and lower complexity transition skills yielded the strongest effects, but transition tasks only included 16 contrasts compared to 31 contrasts for social skill tasks; therefore, these data should be interpreted with caution. Tasks with the lowest complexity (i.e., 1 to 10 steps) included 53 contrasts and yielded a moderate Tau-U score of 0.84 CI_{95} [0.73, 0.95] indicating a more precise measure of the efficacy of video modeling interventions when teaching lower complexity tasks to individuals with. Overall, low complexity tasks and high complexity social tasks yielded moderate to strong effects with narrow confidence intervals and a large number (n = 53, 31) of contrasts per level; therefore, these results can be interpreted with more confidence compared to the small number (n =19, 4, 16) of contrasts for moderate complexity, high complexity, and low complexity transition tasks.

The Kruskal-Wallis did indicate a significant difference between task complexity moderator levels (p = 0.0021). The Dunn post-hoc yielded a pairwise significant difference between the efficacies of video modeling interventions for low versus moderate complexity tasks, moderate versus high complexity social tasks, and moderate versus low complexity transition tasks. Video modeling interventions targeting low

complexity tasks, high complexity social tasks, and low complexity transition tasks had significantly stronger effects compared to their counterparts (see Table C3).

Participant Characteristics

Three moderator variables focusing on specific participant characteristics (i.e., age, IQ, and diagnosis) were analyzed using Tau-U effect sizes for each level. Each variable's number of studies, number of participants, number of analyses, Tau-U scores are reported in Table C4.

Age. Tau-U scores for video modeling interventions with participants differentiated by age group ranged from moderate effects of 0.82 CI_{95} [0.77, 0.88] for participants ages 16 to 21 years old and 0.83 CI_{95} [0.75, 0.90] for those 22 years old or older to strong effects of 0.85 CI_{95} [0.75, 0.96] for those 12 to 15 years old (see Table 4). The Kruskal-Wallis and Dunn post-hoc tests did not show any significant difference between levels ($p = 0.1415$).

IQ. Participants with high (i.e., 50 or above) and low IQs (i.e., below 50) both yielded strong Tau-U scores of 0.87 CI_{95} [0.80, 0.94] and 0.92 CI_{95} [0.85, 0.99] respectively. High and low IQ variable levels included a similar number of contrasts (n = 43, 44), which creates a more precise comparison of effect sizes (see Table C4). The Wilcoxon two-sample test did not show any significant difference between IQ moderator levels ($p = 0.2620$).

Diagnosis. The diagnosis of participants was measured according to they type of developmental disability (i.e., autism spectrum disorder [ASD], intellectual disability [ID]) including comorbid ASD with ID. Tau-U scores for video modeling interventions

that were implemented with participants with specific diagnoses ranged from moderate effects of 0.79 CI_{95} [0.68, 0.89] for participants with ASD and 0.81 CI_{95} [0.73, 0.90] for participants with ASD with ID to strong effects of 0.85 CI_{95} [0.80, 0.91] for participants with ID. Individuals with comorbid ASD with ID and individuals with ID included the largest number of contrasts (n = 45, 57); therefore, the strong effects for individuals with ASD with ID and moderate effects for individuals with ID can be interpreted more precisely compared to individuals with ASD (n = 10 contrasts). The Kruskal-Wallis and Dunn post-hoc tests did not show a significant difference between the participants in each diagnostic category (p = 0.3528).

Setting

The Tau-U effect sizes of video modeling interventions conducted in natural and simulated settings were analyzed individually and comparatively. Video modeling interventions implemented in natural settings yielded moderate effects 0.81 CI_{95} [0.75, 0.87] and interventions implemented in simulated settings yielded strong effects 0.85 CI_{95} [0.79, 0.91]. Both setting moderator levels included more than 40 contrasts each; therefore, more precise conclusions could be made based on outcomes (see Table C5). The Wilcoxon two-sample test did not show any significant difference between setting moderator levels (p = 0.1016).

Interventionist

Analyses of video modeling interventions implemented by three different interventionists were conducted. Tau-U scores for interventionist variables ranged from weak effects 0.65 CI_{95} [0.79, 0.92] for implementation by job coaches, to moderate

effects 0.78 CI_{95} [0.64, 0.92] for implementation by teachers, to strong effects 0.92 CI_{95} [0.79, 1.00] for implementation by researchers (see Table C5). Due to the lack of information given to describe the interventionists in each study, the number of contrasts (n = 12, 14) for job coach and researcher variables were too small to break down effect sizes for each level. Therefore, the Tau-U scores gleaned for each interventionist moderator level are inconclusive.

IRR

IRR for inclusion and exclusion processes, moderator coding, and data extraction were scored for percentage agreement and Cohen's kappa (Cohen, 1960). The title/abstract and full text IRR was scored as 99% and 93% agreement with kappa scores of 0.72 and 0.84 respectively. The basic design standards screening yielded an IRR of 96% agreement and a kappa score of 0.89. Moderator coding IRR was calculated as 94% agreement. Data extraction IRR yielded 97% agreement.

Discussion

This meta-analysis analyzed the outcomes of 39 SCED studies that implemented a range of interventions to promote acquisition and independence in performing employment-related behaviors for individuals with developmental disabilities. This meta-analysis contributes to the literature regarding employment skills and is unique in its focus, statistical methodology, and analysis of relevant variables. The results of this meta-analysis provide evidence of moderate to strong effects across all employment skills interventions. These results corroborate the outcomes of Ninci et al. (2015)

regarding employment skill interventions for a sample of articles that only included individuals with ASD.

The first question considered in this meta-analysis inquired about the difference in effects across intervention types. Strong effects were found for three out of the four interventions (i.e., audio cueing/coaching, visuals, and prompting). Moderate effects resulted from video modeling interventions, which yielded overall lower effect sizes than audio cueing/coaching, visual, and prompting interventions. Despite lower effects, video modeling interventions resulted in moderate effects that are consistent with prior work that also categorized video modeling as a quality, effective, and evidence-based practice for individuals with developmental disabilities (Bellini & Akullian, 2007; Mason, Davis, Boles, & Goodwyn, 2013a; Mason, Ganz, Parker, Burke, & Camargo, 2012a).

There was a pair-wise significant difference in effects between audio cueing/coaching and video modeling interventions according to the Dunn post-hoc results. Audio cueing/coaching interventions yielded stronger effects than video modeling interventions. However, the small sample of audio cueing/coaching studies solely contained studies with highly positive outcomes, which may not be representative of all audio cueing/coaching intervention effects. This difference in effect should be considered with caution due to the small sample size. To attain a more accurate effect size for audio cueing/coaching, researchers need to replicate this employment skill intervention as well as the other promising interventions (i.e., visual and prompting interventions).

The second question considered in this meta-analysis inquired about the difference in video modeling intervention effects across the type and complexity of target behaviors. Video modeling interventions targeted job-specific skills, including (a) repetitive, (b) assembly, and (c) cleaning; and employability skills, including (a) social and (b) transition skills; all of which yielded moderate to strong effects across target behaviors. Many of the behaviors and video models of these behaviors were constructed with thorough task analyses, which may contribute to the success of the participants in accomplishing these tasks. Repetitive behaviors yielded the weakest effects (i.e., 0.69) compared to other behaviors. This may be due to the diversity of tasks included in repetitive behaviors (i.e., copying, faxing, stocking), which may account for the variability of outcomes seen in the high number of contrasts analyzed for these behaviors. There was a pair-wise significant difference in effects between job-specific assembly and repetitive tasks. The stronger effects seen in assembly tasks may be due to less diversity of task complexity (i.e., low to moderate complexity) than repetitive behaviors (i.e., low, moderate, and high complexity), which may have resulted in better participant performance with assembly tasks.

Social skill behaviors yielded strong effects and more than 30 contrasts were analyzed for this skill. This finding may give more insight into the effectiveness of video modeling in teaching a diverse set of social skills (i.e., interacting with customers, interacting with employers, and interacting with others during an interview) and other less structured employability skills. Social skills are complex and very difficult to teach in a variety of contexts; so, more research should be done to verify the importance of

video modeling interventions in teaching social skills. Overall, these strong effects across a diverse set of job specific and employability skills is promising for individualizing interventions based on individuals' needs for employment skill acquisition.

In addition to the effects of interventions targeting certain types of employment skills, the complexity (i.e., 1-10 steps, 11-20 steps, 21 or more steps, complex social skills, or simple transition tasks) of these skills was analyzed for video modeling studies. All levels of task complexity yielded moderate to strong effects. Low (i.e., transition from task to task) and high complexity (i.e., social skills) tasks yielded the strongest effects. Pair-wise significant differences were found between stronger effects of low, social, and transition skill complexity compared to lower effects seen in skills of moderate complexity. Video modeling interventions have been determined by previous studies as effective in teaching individuals with ASD and other developmental disabilities a variety of functional living skills (i.e., independent living, social, and employment skills), which support the moderate to strong effects across the range of skills and task complexities (Mason et al., 2012a, 2013a; Ninci et al., 2015). The significant differences between task complexities may be due to the small number of individuals from each diagnostic category who were taught at each level of task complexity. There were not enough participants in each diagnostic category for each level of complexity to determine if participant diagnosis was related to the acquisition and performance of certain tasks. Also, the significant difference could have resulted in the use of different video modeling components to teach these skills (e.g., video

prompting versus video priming), which could make a difference in the acquisition and performance of designated task analyses. Overall, these results indicate that skills requiring a range of skill levels can be taught, acquired, and performed as a result of video modeling interventions. The range of complexity allows for researchers and practitioners to implement more individualized video modeling interventions for individuals with developmental disabilities with a variety of deficits in employment skills.

The third question considered in this meta-analysis inquired about the difference in video modeling intervention effects across participant characteristics (i.e., age, diagnosis, and IQ). Moderate and strong effects were calculated across participants ages 12 years old and older (see Table C4). A majority of studies implemented video modeling interventions for participants ages 16 to 21 years old; resulting in moderate intervention effects. These moderate effects for individuals 16 years and older corroborates the IDEA (2004) federal mandate to begin transition planning no later than age 16. IDEA (2004) also mandates that transition planning can begin whenever the individualized education program (IEP) team determines it to be appropriate. Video modeling interventions implemented with individuals ages 12 to 15 years old yielded the strongest effects (0.85) compared to other age groups. With many special education stakeholders calling for transition planning to begin before the age of 16, efficacy of video modeling interventions with these younger age groups is promising in teaching employment or pre-employment skills. If transition planning begins at age 12, this leaves up to 9 years of employment skill instruction before these individuals graduate high

school (Johnson, 2012). This may allow for more preparation for individuals to acquire employment and successfully fulfill the duties of employment.

Strong effects were also seen for individuals across IQ scores, and participants diagnosed with ID. No significant differences existed between individuals with ASD, ASD with ID, and ID, or individuals with low and high IQ scores, which indicates moderate to strong video modeling intervention effects can occur for individuals with range of developmental disabilities and intellectual functioning levels (see Table C4). The lack of significant differences within IQ and diagnosis variable levels may be due to the inability to assess detailed similarities and differences between these levels. For example, cognitive functioning is typically related to IQ and adaptive skills, but only a small portion of the included studies reported adaptive assessment scores, making it difficult to assess which component of cognitive functioning may influence intervention effects. More research is needed in implementing employment skill interventions with a diverse population of individuals with DD to assess the detailed components of variables and their influence on intervention effects. Overall, these IQ and diagnosis results mirror previous reviews and meta-analyses outcomes determining the efficacy of video modeling interventions for individuals with ASD and other DD (Bellini & Akullian, 2007; Mason et al., 2012a, 2013a).

The final question considered in this meta-analysis inquired about the difference in video modeling effects across settings and interventionist characteristics. Video modeling interventions implemented in both natural and simulated settings yielded moderate effects in the natural setting and strong effects in the simulated setting with no

significant difference in efficacy. Similar effect sizes for these settings may be due to the fact that simulated settings replicate the materials, the individuals present, and/or the environmental arrangement of the setting in which the employment skill would typically take place in (i.e., natural setting; Lattimore et al., 2006; Test et al., 2012; Walker et al., 2012). These effect sizes are promising for the practicality and efficacy of interventions implemented in a simulated setting if implementation in a natural setting is not plausible. In many cases there are limited resources (i.e., lack of time, facilities, job coaches) when implementing employment skill interventions and in some cases, it is not practical to teach these skills in a busy and over stimulating work environment. These results should be interpreted with caution due to the absence of generalization analyses in this meta-analysis.

In addition to the setting, video modeling intervention effect were inconclusive in analyses for interventionist characteristics. The lack of reporting interventionist characteristics in many studies led to the small number of contrasts for researchers and job coaches. Therefore, future research should focus on thoroughly reporting interventionist characteristics and video modeling implementation by a variety of individuals present in the everyday training or work environment of the individual with a disability (i.e., teachers, job coaches, and co-workers, employers) to promote generalization of skills and maintenance of intervention implementation in the natural environment (Likins, Salzberg, Stowitschek, Lignugaris-Kraft, & Curl, 1989). If research replicates video modeling or other employment skill interventions with a

variety of interventionists, effect size analyses can produce more precise conclusions in the future.

Limitations

Overall, a majority of interventions and moderators yielded strong effects, but there were three limitations in this meta-analysis. First, there were studies or contrasts that included small sample sizes for some variable levels (i.e., participants 12 to 15 years old, diagnosis of ASD, audio cueing/coaching and prompting interventions, researcher interventionists, and transition skills). The small number of studies or contrasts per level may have inflated the strong effect sizes because these samples only included highly positive outcomes, which may not be representative of all intervention or moderator effects. This is also apparent in the presence of larger confidence intervals due to the lack of precise measurement for smaller samples. Small samples in these areas indicate the need for more employment skill research for different interventions, ages, diagnoses, and employment-related skills.

Second, data extraction and effect size analyses did not include maintenance or generalization outcomes. Maintenance and generalization data are important to the continuation of study effects whether in the same or different contexts. Future analyses of these phases are needed to provide more information on the continuation of treatment effects on employment skills.

Third, this meta-analysis focused solely on SCEDs implementing employment skill interventions. Group studies were not considered in these analyses due to the difficulty of comparing two different effects sizes from SCED and group experimental

design. Group studies should be analyzed in future reviews to offer additional insights into overall employment skill intervention effects.

This meta-analysis thoroughly analyzed relevant data and specific moderators to inform researchers and practitioners of employment skill interventions' effectiveness in multiple contexts. Video modeling techniques, in general, have been defined as portable, efficient, and socially acceptable interventions that can be beneficial to adolescent and adult individuals with developmental disabilities across contexts (Mason et al., 2012a, 2013a). This meta-analysis applied measures of efficacy to video modeling interventions that focused on teaching employment skills. Video modeling interventions yielded moderate to strong effects across a diverse group of participants, study elements, and employment skills. These results provide a more solid foundation for researchers and practitioners to individualize and implement effective employment skill interventions.

CHAPTER III

QUALITY REVIEW OF SINGLE-CASE STUDIES CONCERNING EMPLOYMENT SKILL INTERVENTIONS FOR INDIVIDUALS WITH DEVELOPMENTAL DISABILITIES

Many individuals with developmental disabilities (DD) have difficulty transitioning from school to a career due to deficits in communication, social interaction, and task completion (Grigal & Deschamps, 2012; Hendricks, 2010; Hendricks & Wehman, 2009; Wehmeyer, 1994). These are essential skills for attaining and maintaining competitive employment, but employment outcomes of individuals with DD are relatively poor in comparison to their peers without disabilities (Carter et al., 2012; Hanley-Maxwell & Izzo, 2012; Newman et al., 2009, 2011; Sanford et al., 2011). To address this gap in employment outcomes, a variety of curricula and specialized interventions have been incorporated into transition programs to facilitate students' employment after high school (Alwell & Cobb, 2009; Individuals with Disabilities Education Act [IDEA], 2004). Although there has been a recent surge of research focusing on interventions for adolescents and adults with DD, analyses synthesizing this body of research on the effects of interventions to teach and improve employment skills are lacking (Rusch & Datillo, 2012).

Synthesizing bodies of research based on quality is integral to the advancement of researcher and practitioner knowledge of reliable and effective practices (Kratochwill et al., 2013). In the field of education, assessment of the quality of interventions is

guided by federal legislation included in the No Child Left Behind Act (NCLB; 2001) and IDEA (2004). Both NCLB (2001) and IDEA (2004) require and promote the use of scientifically-based research practices, and seek to assess the overall quality and effectiveness of researched interventions. NCLB (2001) outlines key components of scientifically-based research practices: (a) systematic methodological elements in observation or experimental contexts, (b) systematic procedures based on statistics for analyzing data, (c) valid and reliable measures for data collection, (d) study designs that validly measure relations between the intervention and outcomes, (e) thorough descriptions of study characteristics to allow for replication or the growth of future research, and (f) acceptance of the publication through peer-review or more intensive review processes. This 6-component evaluation schema for scientifically-based research is intentionally broad and meant to include a variety of study designs and elements (i.e., group and single-case experimental design [SCED]).

It is important to apply the components of scientifically-based practices, sometimes referred to as evidence-based practices, to SCED in order to assess the quality of applied interventions (Horner et al., 2005; Kratochwill et al., 2013). Thorough quality analyses of SCEDs can instill confidence in quality studies' intervention effectiveness within specified contexts (Kratochwill et al., 2013). SCED continues to contribute to special education reform and the practices geared towards individualized instruction for individuals with disabilities (Gast & Ledford, 2012; Horner et al., 2005). Although SCEDs focus on individual participants, a quality analysis of multiple studies and participants can create a foundation for discerning evidence-based practices (Horner

et al., 2005; Kratochwill et al., 2013). In order to appropriately apply evidence-based practice standards to SCED, an aggregate of studies implementing a specific intervention need to meet the following criteria: (a) at least 5 studies with high-quality designs that exhibit a functional relation between the interventions and target behaviors, (b) at least 3 different research groups (no author repeats) conducted the research at 3 separate institutions; and (c) a combination of at least 20 experiments from the included studies (Horner et al., 2005; Kratochwill et al., 2010, 2013). These three requirements (also referred to as 5-3-20) define the basic foundation of considering evidence-based practices.

In order to deem a study as high quality within the evidence-based practices qualification process, individual quality indicators must be assessed (Horner et al., 2005; Kratochwill et al., 2010, 2013). These quality indicators should be operationally defined to avoid error in consistency of quality ratings across studies (Cooper, 2010). Quality indicators can be applied in stages and *What Works Clearinghouse* (WWC) provides basic design indicators for inclusion or exclusion of possible studies (Kratochwill et al., 2010; Ninci et al., 2015). The basic design standards include: (a) purposeful manipulation of the independent variable (IV); (b) interobserver agreement (IOA) is recorded for 20% of overall data, resulting in an overall score of at least 80% agreement; (c) three different attempts to present effect at three separate points in time; and (d) each phase contains at least 3 to 5 data points. Studies are then categorized as meeting these design standards, meeting these standards with reservations, or not meeting these standards (Kratochwill et al., 2010, 2013). After the exclusion of studies that do not meet

the basic design standards, descriptive indicators can be applied to assess the overall quality of each study without excluding additional studies. These standards include: (a) the possibility of replication based on detail given for participant characteristics, setting characteristics, interventionist characteristics, baseline and intervention procedures, and definition and measurement of dependent variables (CEC, 2014; Horner et al., 2005; Palmen, Didden, & Lang, 2012; Reichow et al., 2008; Roth et al., 2014; Wolery, 2013); and (b) the presence and measurement of generalization, maintenance, procedural fidelity, and social validity data (Banda et al., 2011; CEC, 2014; Reichow et al., 2008; Walker, 2010; Wolery, 2013).

For all studies that meet the minimum basic design standards, visual analysis is a necessary step when analyzing the quality of intervention effects (Kratochwill et al., 2013). Visual analysis is broken up into multiple evidence quality indicators that are then applied to the studies that either meet or meet the design standards with reservations. (Kratochwill et al., 2013; Maggin et al., 2013; Ninci et al., 2015). The evidence standards include the visual analysis of (a) level, (b) trend, (c) variability, (d) immediacy of effect, (e) overlap, and (f) consistency of data patterns in similar phases seen within and between baseline and intervention phases (Kratochwill et al., 2010, 2013). Unfortunately, there are a lack of reviews and meta-analyses incorporating these basic design and evidence standards when assessing the quality of studies focusing on SCED and employment skills for individuals with a range of DD (Ninci et al., 2015; Palmen et al., 2012; Roth et al., 2014; Taylor et al., 2012).

Multiple reviews and meta-analyses have assessed the quality of studies using at least one of these quality indicators listed above (e.g., CEC, 2014; Kratochwill et al., 2013), but these reviews have not combined indicators from a variety of sources to address every quality aspect of a study under investigation (Banda et al., 2011; Ninci et al., 2015; Palmen et al., 2012; Roth et al., 2014; Taylor et al., 2012; Walker et al., 2010). There is a growing need to combine all relevant SCED quality indicators for an overall quality analysis of an entire body of literature for a variety of employment skill interventions focusing on individuals with a range of DDs (Ninci et al., 2015; Palmen et al., 2012; Roth et al., 2014; Taylor et al., 2012; Walker et al., 2010).

Currently, no meta-analyses or reviews have analyzed the quality of research on multiple types of employment skill interventions for individuals with a range of DDs using multiple quality indicator sources. A comprehensive and systematic quality analysis of SCED employment skill studies can inform special education teachers, practitioners, and researchers of promising or evidence-based interventions for individuals with DD.

The purpose of this quality analysis of SCED studies implementing employment interventions for individuals with DDs is to address gaps in the current body of research and provide a response to the following question:

1. Does the body of SCED research on employment skills for adolescent and adult populations with DDs meet minimum design and evidence standards as well as adhere to descriptive design quality indicators (i.e., CEC 2014; Horner et

al., 2005; Kratochwill et al., 2013; Maggin et al., 2013; Reichow et al., 2008; and Wolery, 2013)?

Methods

Article Identification

Search procedures. An electronic database search for potential studies was conducted using an electronic search engine. The databases included: *(a) Academic Search Complete, (b) Applied Technology Full Text, (c) ERIC, (d) Education Full Text, (e) Professional Development Collection*, *(f) Psychology and Behavioral Sciences Collection, (g) Social Science Full Text, (h) Vocational and Career Collection,* and *(i) Vocational Studies Complete*. All peer-reviewed and non peer-reviewed sources were retrieved from these databases. Two groups of search terms were used as Boolean phrases (includes the word *and* in between the key search terms) when searching the databases. The first group of terms included: *autis*, Asperger*, ASD, PDD*, pervasive developmental disorder, development* disab*, low-incidence dis*, intellectual* disab*, mental* retard*,* or *multiple disab**. The second group of terms included: *employ*, career*, vocation*, employ* skill*, career skill**, *vocation* skill*,* or *job skill**. The search terms identified with an asterisk broaden the database search by including the stem of the word and any possible suffix.

Inclusion and Exclusion Criteria

Title and abstract inclusion/exclusion. The title and abstract of each retrieved source were screened using the following criteria: (a) employed a SCED, (b) included at least one participant diagnosed with DD, (c) contained one or more dependent variables

that measured transition skills (i.e., employment skills, independent living skills, social skills), (d) included an intervention component as the IV, (e) reflected a journal article or dissertation, and (f) published in English. Due to the focus of this meta-analysis on employment skill interventions for individuals with DD, documents were excluded if the targeted diagnoses, IVs, and dependent variables criteria were not met. Further, it is important to search peer-reviewed and other (e.g. dissertations or theses) sources to avoid publication bias (Lipsey & Wilson, 2001). If there was insufficient information in the title or abstract to evaluate all inclusion and exclusion criteria, the full text of that document was reviewed.

Full text inclusion/exclusion. Following the title and abstract screening, the remaining articles were evaluated using the full-text. The full text of each article was screened using the following criteria: (a) employed a SCED (i.e., reversal/withdrawal, alternating treatments, multiple baseline, multiple probe, multi element design); (b) contained one or more dependent variables that measured employment skills; (c) included at least one participant diagnosed with DD; (d) implemented an intervention that focused on teaching and promoting independent performance of employment skills; and (e) contained a line-graph representing skill acquisition or independent task performance data (i.e., percent of task steps performed correctly and independently or number of prompts needed to complete a task). The inclusion of a line-graph representation of data was chosen as a criterion due to the need for visual analysis of the data for the quality of design and evidence reviews.

In an effort to identify all available articles pertaining to employment skill interventions for individuals with DD, an ancestral search was also conducted. This entailed searching through the references of previously included studies. Each reference was screened based on the title, following the earlier described procedures; those references determined to reflect potential studies for inclusion were pulled and the full-text was evaluated using the full set of inclusion and exclusion criteria.

Interrater reliability (IRR). Two raters independently screened 30% of the sources retrieved after the database search for the title and abstract and full text inclusion and exclusion screenings. The articles for interrater review were randomly selected from the total number of sources to avoid selection bias. In the case of a disagreement in any of the inclusion and exclusion processes, the two raters discussed the discrepancy and reached a consensus without the need of a third rater. IRR was scored as simple percent agreement (total number of agreements divided by agreements plus disagreements and then multiplied by 100) and Cohen's kappa (Cohen, 1960).

Design Quality Indicators

Basic Design Standards. Documents that passed title/abstract and full text reviews were further reviewed to whether or not each experiment present in the study met the minimum design standards (Kratochwill et al., 2010, 2013). Experiments are defined as each data representation of a single-case design (Kratochwill et al., 2010). For example, if there were two participants in a study and a multiple baseline design across skills was conducted for each participant; two different experimental data sets were independently screened according to the basic design standards. The design standards

included the following criteria: (a) purposeful manipulation of the IV; (b) IOA recorded for 20% of overall data, resulting in an overall score of at least 80% agreement (IOA components are broken up into three individually rated standards); (c) three different attempts to demonstrate an effect at three separate points in time; and (d) each phase contained at least 3 data points. Systematic manipulation of the IV is important when assessing the functional relation by applying certain conditions purposefully (Kratochwill et al., 2013). IRR/IOA represents the agreement between two raters or observers when collecting data, which provides a measure for reliability (Horner et al., 2005; Kratochwill et al., 2013); therefore, IRR/IOA needs to be measured often and across time to ensure consistent reliability of the measures**.** Demonstrating or attempting to demonstrate an effect is necessary to assessing the consistent functional relation across time and should occur at least three times over three different time periods (Kratochwill et al., 2013). The number of data points is important when assessing the consistency of data and if behaviors are really changing from baseline to intervention. Five data points is preferred because working with individuals usually creates natural variability in the data, which can make it difficult to assess consistency if there are less than 5 data points per phase (Kratochwill et al., 2013).

Each experiment was analyzed according to each design standard using the rating system presented and defined in Boles (2015a, available online). Dichotomous ratings of 0 (i.e., does not meet design standards) or 2 (i.e., meets design standards) were used to assess the purposeful manipulation of the IV and attempts to present an effect. A 3 item rating system including the scores 0, 1 (i.e., meets design standards with reservations),

and 2 was used for IRR measures and the number of data points because these two standards have an accepted quality measure (i.e. 3 to 4 data points per phase) and a preferred quality measure (i.e. 5 or more data points per phase).

After each study was scored according to the design standards, an overall score was assigned to each study as a whole. If a study contained at least one experiment that met or met with reservations the design standards, the entire study was scored based on this experiment. An overall score of 0 (i.e., does not meet the overall design standards) was given if one or more of the design standards listed above were scored with a zero. An overall score of 1 (i.e., meets overall design standards with reservations) was given if at least one of the design standards listed above was scored with a 1 and all the other standards were scored as 1 or 2. Finally, an overall score of 2 (i.e., meets overall design standards) was only given if all design standards were scored with a 2.

IRR for basic design standards. The basic design standards screening was completed by two raters for 50% of the articles remaining after the title/abstract and full text inclusion and exclusion process. In the case of a disagreement when evaluating the basic design standards, the two raters discussed the discrepancy and reached a consensus. IRR was scored as simple percent agreement and Cohen's kappa (Cohen, 1960).

Descriptive Design Quality Indicators

In addition to the application of basic design standards, there were key descriptive design quality indicators applied to all studies that met or met with reservations the basic design standards (CEC, 2014; Kratochwill et al., 2010, 2013; Horner et al., 2005; Reichow et al., 2008; Wolery, 2013). Descriptive design indicators

are those indicators that are rated according to the description and measures of all relevant elements included in each study. The description of specific study characteristics was important to analyze because sufficient detail fosters accurate replication in future research. Replication is crucial in corroborating study effects and strengthening external validity (Horner et al., 2005). In this analysis, there were five indicators that were rated according to the level of replicability based on descriptive detail: (a) participant description, (b) setting description, (c) interventionist description, (d) baseline and intervention description, and (e) dependent variable description. This rating scale along with an overall score (i.e., *Insufficient Description*, *Minimal Description*, and *Sufficient Description*) according to the level of replicability is described in Boles (2015b, available online). In addition to the five replicability indicators, four additional descriptive design indicators based on supplementary measures or assessments in each study were included: (a) maintenance, (b) generalization, (c) fidelity, and (d) social validity. The rating system for these four indicators and overall scores (i.e., *Insufficient Measure*, *Minimal Measure*, and *Sufficient Measure*) is described in Boles (2015c, available online). All of these indicators were purely descriptive with overall quality scores, and studies were not excluded based on descriptive design indicator scores.

Participant Description. In each SCED study, the participant should be described in enough detail to promote replication of the population being targeted in future research. An operational definition of a participant should include the specific diagnosis and the assessments or process that lead to the diagnosis (Horner et al., 2005).

Also, participants should be identified by age, gender, and any other relevant characteristic (i.e., IQ, skill deficits, previous training/therapy; CEC, 2014; Reichow et al., 2008). Participant descriptions were measured using a rating scale found in Boles (2015b).

Setting Description. A thorough description of the setting is an important element to consider when replicating a study (Horner et al., 2005). Setting elements such as the materials and layout of the setting, the presence of other individuals, and the location (i.e., classroom, home, work) can impact the effects of the intervention (CEC, 2014). The setting description was measured using a rating scale found in Boles (2015b).

Interventionist Description. The characteristics of the interventionist are necessary when measuring the effectiveness of intervention implementation based on the interventionist's expertise and relationship towards the individual receiving the intervention. The interventionist description should include the interventionist's occupation and relationship to the participant (i.e., teacher, peer, sibling, parent, researcher), and the interventionist's level of expertise in implementing the intervention (CEC, 2014). The interventionist description was measured using a rating scale found in Boles (2015b).

Baseline/Intervention procedure description. The baseline and intervention procedures are necessary when assessing the steps taken to prepare for and implement an intervention. A thorough description of these procedures is necessary for accurate replication and reliable measures (Horner et al., 2005). Baseline and intervention descriptions should include a thorough description of the baseline procedures (i.e.,

setting, materials used, assessed behaviors, session time limit) and intervention procedures (i.e., chronological steps for implementing the intervention, the behaviors required of the interventionist, setting, materials used, session time limit; Horner et al., 2005; Reichow et al., 2008). The baseline/intervention procedure description was measured using a rating scale found in Boles (2015b).

Dependent variable description. The dependent variable is important in its role in determining the success of the intervention and the overall functional relation between the intervention and the targeted behaviors. An operational definition of the dependent variables is needed to promote a valid, reliable, and objective measure of scientific observation (Horner et al., 2005). The operational definition of target behaviors, the reasons for targeting these behaviors, and a thorough description of data collection methods for the targeted behaviors are needed for future replication (CEC, 2014; Horner, et al., 2005; Reichow et al., 2008). The dependent variable description was measured using a rating scale found in Boles (2015b).

Maintenance and generalization. Maintenance and generalization enhance external validity by providing long-term data collection and/or data collection in multiple contexts (e.g., different materials, participants, interventionists, settings; Horner et al., 2005). Studies may or may not include maintenance and generalization data, but both are important to study quality and the ongoing effects of an intervention. Maintenance is measured by assessing the progress of target skills over time either with continued implementation of the intervention or as a result of the withdrawal of the intervention (Horner et al., 2005; Kazdin, 2011). In the case of employment skill

interventions, it is very important to record maintenance data due to the goal of not only acquiring but also maintaining employment skills. Generalization is considered the measure of certain effects in novel or different contexts that may include multiple participants, settings, materials, or interventionists (Horner et al., 2005). Generalization is important when assessing quality because those participants receiving an employment intervention need to know how to apply newly acquired skills to different contexts that may occur during the transition into employment (Horner et al., 2005; Reichow et al., 2008). The rating scale and description of maintenance and generalization indicators can be found in Boles (2015c).

Fidelity. Fidelity is not always included as a measure of study quality, but these are important measures when assessing the accuracy and consistency of implementation. Procedural or treatment fidelity measures the accuracy or human error when implementing the procedures included in all conditions or only in the intervention phases (Ledford & Wolery, 2013). Errors in fidelity can weaken internal validity due to the intervention being implemented over time when maturation and other variables can play a role in behavior change outside of the results of the intervention (Horner et al., 2005; Ledford & Wolery, 2013; Wolery, 2013). Procedural or treatment fidelity measures should be recorded throughout the intervention or all phases using a form of data collection that measures accuracy of implementation by the interventionist for each step included in the procedures (CEC, 2014; Horner et al., 2005; and Reichow et al., 2008). The rating scale and description of the fidelity indicator can be found in Boles (2015c).

Social Validity. Social validity is defined as the overall acceptability of the procedures and outcome measures involved in an intervention program (Carter, 2010). Social validity is crucial in maintaining an intervention program and the effects of that program (Carter, 2010; Schwartz & Baer, 1991). Social validity should measure the (a) social significance of the dependent variables (i.e., the target behaviors are beneficial to the participant and relevant to the context), (b) the efficiency and cost effectiveness of the intervention, (c) the significance of behavior change or intervention effects were significant according to the criteria or goals set for individual studies, (d) the satisfaction of all individuals involved regarding the procedures and outcomes, and (e) the inclusion of a natural component in the intervention (i.e., the interventionist is an individual that is present in the participant's natural setting, or the intervention is implemented in the natural setting; Horner et al., 2005; and Reichow et al., 2008; Wolery, 1978). The rating scale and description of the social validity indicator can be found in Boles (2015c).

IRR for descriptive indicators. The descriptive quality indicator analysis was completed by two raters for 100% of the articles remaining after basic design standards inclusion and exclusion process. In the case of a disagreement when evaluating the quality indicators the two raters discussed the discrepancy and reached a consensus. IRR was scored as simple percent agreement and Cohen's kappa (Cohen, 1960).

Evidence Quality Standards

Visual analysis is crucial when analyzing the overall quality of reported effects in SCED studies (Kratochwill et al., 2013). It is recommended that visual analysis be conducted when assessing evidence (Brossart, Vannest, Davis, & Patience, 2014;

Kratochwill et al., 2013). Visual analysis requires the review of the main components of each experiment: (a) level, (b) trend, (c) variability, (d) immediacy of effect, (e) overlap, and (f) consistency of data patterns in similar phases seen within and between baseline and intervention phases. Level is defined as the average measure of each phase. Variability takes into account the overall consistency or inconsistency of data throughout each phase. Immediacy of effect relies on the level of the last three data points in baseline compared to the level of the first three data points in the intervention phase. Finally, the consistency of data in similar phases was analyzed based on the similarity between the level, trend, and variability seen in data sets present in similar phases (i.e., comparison of data consistency in baseline phases, A_1 and A_2 of a reversal design [ABAB]; Kratochwill et al., 2013). These six visual analysis components were applied to four different evidence indicators (a total of 19 different items): (a) within-phase data points, (b) overall data points, (c) overall ratio of effects to non-effects, and (d) overall evidence of effect. A rating system found in Boles (2015d) for each component of visual analysis of evidence was applied to each experiment from included studies, resulting in studies categorized as visually presenting *No Evidence*, *Moderate Evidence*, or *Strong Evidence* (Kratochwill et al., 2010, 2013; Maggin et al., 2013).

In order to analyze the experiments and overall studies for quality of evidence and possible declaration of an evidence-based practice, categorization via the type of primary intervention implemented in each experiment needs to occur. Intervention codes for each experiment were employed: (a) VM (i.e., video modeling; the use of a peer, adult, participant, or point-of-view perspective in modeling target behaviors as a video

or in-vivo presentation for a variety of implementations: video modeling, video prompting, video priming, video self-modeling, point of view video modeling, adult/peer video modeling, or in-vivo modeling), (b) AC (i.e., audio cueing or audio coaching delivered to the participant via an earpiece or other device while the task is performed), (c) VIS (i.e., any static pictures, written schedules, picture schedules, or scripts that prompt a participant through a task), (d) PRMTS (i.e., most-to-least or least-to-most prompting, or any other systematic prompting system using least or most intrusive prompts as the primary intervention), (e) OTH (i.e., any intervention that does not fit the categories above or combines more than one of the specified interventions). If the experiments were scored with moderate or strong evidence, they were included in the evidence-based practice analysis based on the intervention employed and the 5-3-20 evidence-based rule.

IRR for IV codes and evidence standards. The IV coding and evidence standards analysis were completed by two raters for 100% of the articles remaining after the basic design and evidence standards inclusion and exclusion process. In the case of a disagreement when evaluating the IV codes or evidence standards, the two raters discussed the discrepancy and reached a consensus. IRR was scored as simple percent agreement and Cohen's kappa (Cohen, 1960).

Results

The overall article search from designated databases resulted in 5,821 possible articles with the removal of duplicates. These articles were analyzed using the title and abstract and full text inclusion and exclusion criteria stated above. The abstract inclusion

and exclusion screenings resulted in 240 articles and full text screenings resulted in 79 articles. The basic design standards (Kratochwill et al., 2010, 2013) were applied to these remaining articles and resulted in 34 articles that passed all inclusion and exclusion criteria. An ancestral search of the reference section in each of the 34 articles was performed in order to find any articles that were not included in the initial search due to the search criteria or human error. Forty-three additional articles were found during the ancestral search and these were screened based on the abstract and title, full text, and basic design standards criteria. These screenings resulted in 2 additional studies for inclusion in the meta-analysis. A total of 36 articles (39 separate studies) were analyzed using the quality indicators described above. Tables D1 to D4 in Appendix D provide the final analysis of each included study based on the ratings derived from the basic design standards, descriptive design quality indicators, and the evidence quality standards indicated in Boles (2015a, 2015b, 2015c, 2015d).

IRR for overall search. The IRR agreement for the abstract and title screenings was calculated as 99% with a kappa score of 0.72. IRR agreement for the full text screening was 93% with a kappa score of 0.84. Lastly, the IRR agreement for the basic design standards screening was calculated as 96% with a kappa score of 0.89.

Basic Design Standards

As a result of the initial search and the ancestral search, a total of 89 articles were analyzed using the basic design standards and rated according to the scoring system provided by Boles (2015a), based on Kratochwill and colleagues (2010, 2013). The individual experiments that did not meet the design standards or meet them with

reservations recorded IRR for less than 20% of sessions, less than 3 demonstrations of possible effect, and/or less than 3 data points in at least one phase. Table D1 presents a total of 39 studies that passed the basic design standard screening by meeting all design standards or meeting the design standards with reservations. Only 6 of the original 39 studies thoroughly met all design standards. The majority of studies (n = 33) met design standards with reservations. These 33 studies only partially met standards due to reports of IOA session totals, IOA percentage agreement, and/or the number of data points. Twenty-five out of the thirty-three studies reported 20% overall percentage of sessions in which IOA was recorded, but there was no indication of the percentage of sessions for which IOA was recorded per phase or per participant/behavior. Specificity of IOA percent agreement in each phase and each participant/behavior were also missing in 24 of the 33 studies. Lastly, 21 of the 33 studies reported only 3 to 4 data points in at least one phase instead of the preferred 5 or more data points per phase.

Descriptive Design Quality Indicators

The 39 included studies were then analyzed according to the overall descriptive nature of each study design element described in Boles (2015b, 2015c). Table D2 provides the descriptive design quality scores for the participant, setting, interventionist, procedure, and dependent variable descriptions. Table D3 provides the quality scores for the maintenance and generalization phases and the fidelity and social validity measures.

Participant, setting, and interventionist descriptions. A majority of the studies were thorough when providing participant descriptions. Thirty studies provided the participant inclusion criteria, age, gender, primary and secondary diagnoses (if

applicable), IQ scores, and current skill levels or prior therapy. Eight studies partially met the participant description standard by giving broader or less detail (e.g., age range instead of individual ages). Only one study did not meet the standards for participant description because each participant's gender was not reported resulting in a score of 0 (see Table D2).

In contrast to participant descriptions, a majority of studies were not as thorough when reporting the setting description. Only 11 studies provided a thorough description of the setting that included the location, materials present, and presence or absence of other individuals (related or non-related to the study). In 17 studies, the setting was only partially described by including the location and the presence of other individuals or the materials present. The setting was not sufficiently described in 11 studies because only one descriptive element (i.e., location, individuals present, or materials present) was reported.

The interventionist description indicator was similar to the setting description in that a majority of studies received low ratings. In 21 studies, the interventionist in the study was described with a title (e.g., teacher, supervisor, trainer), but the interventionist's expertise (e.g., number of years as a teacher or prior experience implementing the intervention) was never given. In contrast, 11 studies described both important aspects of the interventionist (i.e., occupation/title and expertise). Seven studies did not provide either of these main interventionist descriptions and were therefore awarded a 0 on the rating scale (see Table D2).

Procedure and dependent variable descriptions. Procedure and dependent variable descriptions were reported by a majority of studies with more detail than the setting and interventionist descriptions above. The description of the procedures in 23 studies included an explanation of all necessary elements for both the baseline and intervention procedures (i.e., setting, materials used, session time limit, steps for implementation, and behaviors of the interventionists) with enough detail for accurate replication. In contrast, 15 studies provided only enough replicable details for either the baseline or intervention phase, did not include any indication of session length, or did not include interventionist behaviors. Only one study did not give sufficient detail for either the baseline or intervention phases, resulting in a score of 0 (see Table D2).

The dependent variable description required thorough operational definitions of the target behaviors (i.e., task analysis or detailed description of the task), the reason for targeting specified behaviors, and data collection procedures. Twenty-four studies thoroughly met the dependent variable description standard. In contrast, 14 studies partially met this standard by only reporting either sufficient operational definitions of target behaviors or providing a thorough description of data collection procedures. Only one study did not operationally define the target behaviors or give sufficient detail for data collection procedures, resulting in a score of 0 (see Table D2).

Overall scores. Each of the above quality indicators was taken into account when assigning an overall rating for each of the 39 studies. For the two studies that fully met all descriptive standards, an overall score of 2 or *Sufficient Description* was given. A total of 21 studies met or partially met all descriptive quality standards and were given

an overall score of 1 or *Minimal Description*. A score of 0 for any of the quality indicators above resulted in an overall score of 0 or *Insufficient Description* for 16 studies (see Table D2).

Maintenance and generalization. All 39 studies were analyzed according to the presence of maintenance or generalization data and the quality of these measures (see Table D3). More studies implemented and reported a maintenance phase than those that implemented or reported a generalization phase. Five studies not only reported maintenance measures, but the maintenance phases included data represented by 3 or more data points and recorded more than one month after the conclusion of the intervention. In contrast, 26 studies reported maintenance data, but the data was recorded a month or less after the conclusion of the intervention and/or there were less than 3 data points in this phase. A total of 8 studies did not report any maintenance measures.

Only 5 studies reported generalization measures that occurred in both baseline and intervention sessions with a total of 3 or more data points throughout all generalization measures. In contrast, 16 reported generalization measures, but generalization data was only recorded after the intervention or there were less than 3 total data points for all generalization data. A total of 18 studies did not report any generalization measures.

Fidelity and social validity. Treatment or procedural fidelity and social validity were assessed for each of the 39 studies (see Table D3). Fidelity was reported in 22 studies. Specifically, 14 studies not only reported fidelity measures, but reported fidelity for at least 20% of overall sessions with scores of at least 80% across both baseline and

intervention sessions. In contrast, 9 studies reported fidelity measures for at least 20% of sessions, but fidelity was not recorded in both baseline and intervention phases. Fidelity measures were not reported in 16 studies.

Social validity was the least reported measure compared to maintenance, generalization, and fidelity measures (see Table D3). Social validity was either not reported or only included one element of the five necessary elements to a social validity measure in 21 studies (see Boles 2015c and Table D3). Fourteen studies only reported between two and three of the five necessary elements for social validity. A majority of the studies only reported the social significance of the target behaviors and the significance of the change in behavior according to the goals/criteria set (see Boles, 2015c). In contrast, 4 studies reported at least four of the five necessary elements for a sufficient description of social validity.

Overall scores. Each of the quality indicators above was taken into account when assigning an overall rating for each of the 39 studies. None of the studies fully met all standards for a score of 2 or *Sufficient Measure*. In contrast, 17 studies did meet or partially met a majority of the quality indicators resulting in a score of 1 or *Minimal Measure*. A total of 22 studies did not meet a majority of the standards for a score of 0 or *Insufficient Measure.*

IRR for descriptive quality standards. The overall descriptive quality analysis for participant, setting, interventionist, procedure, and dependent variable descriptions yielded 72% IRR agreement with a kappa score of 0.65. The overall descriptive quality analysis for maintenance, generalization, fidelity, and social validity measures yielded

79% IRR agreement with a kappa score of 0.71. IRR scores for each quality indicator can be found in Appendix D (see Table D5).

Evidence Quality Standards

A total of 83 experiments (i.e., all single-case design data representations present in each article) were analyzed using the evidence quality standards (Kratochwill et al., 2010; 2013) found in Boles (2015d). Four indicators that included a total of 19 different categories were applied to the baseline phase, the intervention phase, the relation between the baseline and intervention phases, and the experiment's overall effects. Table D4 provides the evidence standard scores for all experiments that scored as *Moderate* or *Strong Evidence*. Eight experiments that scored *No Evidence* were excluded from Table D4 because these experiments did not pass the evidence standard screening. Also, unclassified interventions that did not fit the mold of the four primary interventions or included a combination of more than one of the primary interventions (9 experiments; 7 studies) were excluded from Table D4 because the 5-3-20 evidence-based rule could not be applied to these ambiguous interventions. A total of 66 experiments were analyzed according to the 5-3-20 evidence-based rule (Horner et al., 2005; Kratochwill et al., 2010, 2013).

Video modeling. The evidence standard screening for video modeling interventions included 43 experiments (15 studies). All participants in the video modeling studies were 12 years old or older and had a diagnosis of DD. Video modeling studies included 20 participants with ASD, 10 participants with ASD and ID, and 14 participants with ID. A majority of participants (n = 23) ranged from ages 16 to 21 years

old. For the evidence standards, in *Indicator #1*, a majority of the video modeling baseline phases were given the highest scores for the categories: (a) baseline data indicating a participant's need for an intervention for the targeted skill (n = 40 experiments), (b) baseline data indicating predictability (n = 37 experiments), (c) baseline data indicating stability (n = 35 experiments), and (d) baseline data indicating trend toward the hypothesized effect (n = 27 experiments). In contrast, the baseline trend standard included the largest number of experiments (n = 15 experiments) that received a score of 0 (i.e., the baseline trend moves in the opposite direction of the hypothesized direction).

In *Indicator #2*, the experiments' intervention phases were scored based on the number of data points in each phase, data predictability, data variability, and data trend. Scores of 0 or 1 (in the case of a 3 item rating system for the number of data points) were found most commonly in the categories of the number of data points present in each phase (i.e., 3 to 4 data points per phase; n = 37 experiments) and of the consistency of data variability (i.e., intervention phase data fluctuated too erratically to indicate consistency; n = 15 experiments). Only 6 video modeling experiments contained 5 or more data points per phase were given scores of 2.

In *Indicator #3*, the relation between the baseline and intervention phases was scored based on the basic effects between phases, immediate change in level, immediate change in trend, the overall change in level, the overall change in variability, the overall overlap of data between phases, and similarity in data phases (only applicable fore reversal designs). Scores of 0 were most commonly applied in the evidence standard

categories for immediacy of change in level (i.e., no immediate change in level within the first 3 data points of the intervention phase compared to the last 3 data points of the baseline phase; *n* = 17 experiments) and the immediacy of change in trend (i.e., no change in trend within the first 3 intervention data points because the baseline phase already presented a trend toward the intervention's hypothesized direction or the variability of the data made it difficult to visually establish a trend; *n* = 19 experiments).

In *Indictor* #4, the overall evidence of effect was scored as a result of the overall number of data points, overall number of treatment effects, and overall treatment effect ratio. A majority of video modeling experiments received a score of *Moderate Evidence* (n = 30 experiments), and only 6 experiments received a score of *Strong Evidence.* In contrast, 7 experiments received a score of *No Evidence* (i.e., less than 3 treatment effects and/or less than a 3:1 ratio of treatment effects to non-effects) and excluded from the evidence-based analysis.

Evidence-based analysis. As a result of the evidence analysis, 36 video modeling experiments (13 studies) passed the evidence standards and were analyzed according to the 5-3-20 evidence-based rule (Horner et al., 2005; Kratochwill et al., 2010, 2013). Overall, video modeling interventions were implemented by 6 different research groups from separate institutions, and included more than 20 experiments, which indicates that video modeling can be considered an evidence-based intervention in teaching employment skills to individuals with DD.

Audio cueing/coaching. The evidence standard screening for audio cueing/coaching interventions included 5 experiments (4 studies). Audio

cueing/coaching intervention studies included 7 participants with ASD, 2 participants with ASD with ID, and 3 participants with ID; and a majority of participants (*n* = 8) ranged in age from 16 to 21 years old. For the evidence standards, in *Indicator #1*, a total of 3 audio cueing/coaching experiments (3 studies), scored a 0 due to baseline data trend. For *Indicator #2*, scores of 0 or 1 were found most commonly in the categories of the number of data points present in each phase and of the consistency of data variability. A total of 2 audio cueing/coaching experiments (2 studies) received a score of 1 and 3 audio cueing/coaching experiments (2 studies) received a score of 2 for number of intervention phase data points. In addition, 1 audio cueing/coaching experiment (1 study) was scored as 0 due to variability of the intervention phase data. For *Indicator #3*, a score of 0 was most commonly applied in the evidence standard categories of immediacy of change in level (*n* = 1 experiment) and the immediacy of change in trend (*n* = 2 experiments). For *Indictor* #4, the overall evidence of effect for audio cueing/coaching interventions was scored as 1 (*Moderate Evidence*) for 2 experiments (2 studies) and 2 (*Strong Evidence*) for 3 experiments (2 studies).

Evidence-based analysis. As a result of these overall scores, a total of 5 quality audio cueing/coaching experiments (4 studies) passed the evidence standard screening. The audio cueing/coaching interventions were not determined as evidence-based interventions due to the small number of studies/experiments and overlapping author groups (Horner et al., 2005; Kratochwill et al., 2010, 2013).

Visuals. The evidence standard screening for visual interventions included 17 experiments (8 studies). Visual intervention studies included 4 participants with ASD, 0

participants with ASD with ID, and 25 participants with ID. A majority of these participants (n = 23) ranged in age from 16 to 21 years old. For the evidence standards, in *Indicator #1*, a total of 7 visual experiments (6 studies) scored a 0 due to baseline trend. For *Indicator #2*, scores of 0 were applied most commonly in the category of consistency of data variability (n = 3 experiments). Scores of 2 were given to 14 experiments (10 studies) with 5 or more data points per phase. For *Indicator #3*, scores of 0 were most commonly found in the evidence standard categories of immediacy of change in level (n = 4 experiments) and the immediacy of change in trend (n = 5 experiments). In addition, 3 experiments received a score of 1 for the similarity of data patterns in similar phases (i.e., reversal designs). For *Indictor #4*, the overall evidence of effect for visual interventions was scored as 0 (*No Evidence)* for 1 experiment, 1 (*Moderate Evidence*) for 4 experiments (2 studies), and 2 (*Strong Evidence*) for 12 experiments (7 studies).

Evidence-based analysis. As a result of these overall scores, a total of 16 quality visual experiments (8 studies) passed the evidence standard screening. Visual interventions did not meet all of the 5-3-20 evidence-based standards (Horner et al., 2005; Kratochwill et al., 2010, 2013). Overall, there were 8 visual intervention studies implemented by 5 different research groups from separate institutions, but there were less than the required 20 experiments (Kratochwill et al., 2010, 2013), which indicates that visuals cannot be considered an evidence-based practice for employment skills for individuals with DD.

Prompting. The evidence standard screening for prompting interventions included 8 experiments (6 studies). Prompting interventions included 4 participants with ASD, 7 participants with ASD with ID, and 2 participants with ID. A majority of these participants (n = 12) were 22 years old or older. For the evidence standards, in *Indicator #1*, a majority of the prompting baseline phases received a score of 1 in the categories of the need for behavior change (n = 8 experiments), data predictability (n = 7 experiments), and data consistency (n = 6 experiments). In contrast, a total of 6 prompting experiments (4 studies) received a score of 0 due to baseline trend. For *Indicator #2*, scores of 1 were found most commonly for number of data points per phase (n = 6 experiments). In contrast, 2 prompting experiments (2 studies) contained 5 or more data points per phase and received a score of 2. For *Indicator #3*, a score of 0 was only given to one experiment across the evidence standard categories of immediacy of change in level and the immediacy of change in trend. Only one experiment was analyzed according to the category ratings of similar data phases and received a score of 0 due to noticeable difference in data of similar phases. In *Indictor* #4, the overall evidence of effect for prompting interventions was scored as 1 (*Moderate Evidence*) for 2 experiments (2 studies) and a 2 (*Strong Evidence*) for 6 experiments (5 studies).

Evidence-based analysis. As a result of these overall scores, a total of 8 quality, prompting experiments (6 studies) passed the evidence standard screening. The prompting interventions were not determined to be evidence-based interventions due to the overlapping author groups resulting in only two different author research groups/institutions out of the 6 total studies.

IRR for IV codes and evidence standards. The overall IRR agreement score for IV coding was 98% with a kappa score of 0.97. The overall IRR agreement score for evidence standard ratings was 91% with a kappa score of 0.81. IRR scores for each evidence indicator can be found in the Appendix D (Table D5). The majority of disagreements occurred when scoring baseline data variability, baseline data trend, intervention data variability, intervention data trend, the immediacy of change in trend between the baseline and intervention phases, and the overall change in variability between phases.

Discussion

This review analyzed the quality of 79 studies implementing interventions for individuals with DD to promote independence and acquisition of a range of employment skills via the basic design standard screening. A total of 39 studies from the original 79 studies passed the basic design standard screening and were analyzed via the descriptive quality indicators and evidence standards (CEC, 2014; Horner et al., 2005; Kratochwill et al., 2010, 2013; Maggin et al., 2013; Reichow et al., 2008; and Wolery, 2013). Following the exclusion of studies via the evidence standards (Kratochwill et al., 2010, 2013), 38 studies were then categorized according to primary intervention implementation.

The basic design standard analysis resulted in the exclusion of 40 complete studies. The majority of these studies were excluded due to failure to meet 20% IOA across sessions, at least 3 attempts to present effect, and/or at least 3 data points per phase. The year of publication for these studies may play a role in the failure to meet the

basic design standards. Thirty of these forty excluded studies were published earlier than 2005. This may be significant because Horner and colleagues (2005) published research design standards in SCED by calling for better IOA measures, experimental control via 3 attempts to demonstrate effects at 3 different points in time, and 3 or more data points per phase. Many studies published before 2005 may not fulfill the accepted design standards because these standards were not established at that time. Only 6 of the 39 studies met all design standards while 33 studies met the standards with reservations. A majority of the 33 studies lacked specificity when describing IOA procedures and/or only reported 3 to 4 data points per phase. Many studies do not report the percentages of recorded IOA sessions for each participant/behavior or per phase. Overall session percentages are given, which may lead the reader astray when considering just how many sessions per participant/behavior and phase that IOA observations actually resulted. IOA is important when considering the reliability of the data; therefore, it should affect how readers draw conclusions from the overall results of the study (Horner et al., 2005). The preferred minimum of 5 data points per phase is also important as it more clearly represents the predictability, consistency, and trend of the data set (Horner et al., Kratochwill et al., 2010, 2013).

Standards based on the descriptive nature of each of the main elements in each study were analyzed based on a combination of indicators gleaned from CEC (2014), Horner et al. (2005), Reichow et al. (2008), and Wolery (2013). A very small number of studies met the standards for all indicators addressing the description of study elements. The majority of the studies gave insufficient descriptions for the setting or

interventionist. Each element of the study is important to describe thoroughly in order to promote consistent reliability and replication for future researchers who wish to implement a similar intervention and expand the literature (Horner et al., 2005). The only way to build the evidence base for SCED is to promote replicability across authors, institutions, and participants; therefore, all relevant details regarding participant characteristics, settings, interventionists, baseline and intervention procedures, and dependent variables must be at the highest standard to promote replicability (Horner et al., 2005). Further, descriptive characteristics, procedures, and outcomes can inform practitioners of effective interventions that are suited best for certain populations and precise step-by-step procedures for implementing these interventions with fidelity (Reichow et al., 2008).

Beyond baseline and intervention data, detailed and valid measures of maintenance, generalization, procedural fidelity, and social validity are needed to promote replicability and efficacy of specified interventions beyond the confines of the experimental context (CEC, 2014; Horner et al., 2005; Kratochwill et al., 2013; Reichow et al., 2008; Wolery, 2013). The quality analysis for maintenance, generalization, fidelity, and social validity verified that many studies failed to report these measures or only partially met the descriptive standards for these additional measures. None of the studies met all of the quality standards for each measure (i.e., *Sufficient Measure*); therefore, they received overall scores of *Minimal Measure* or *Insufficient Measure*. Generalization, fidelity, and social validity measures were frequently left out or insufficiently described when describing study procedures or results. This is common due to the lack of time or

resources to implement generalization sessions, implement fidelity measures by an observer other than the interventionist, disperse social validity questionnaires to all stakeholders, and/or create a valid social validity measure. Generalization measures are necessary to assess the performance of target skills in a variety of contexts. To validly measure generalization, data recording needs to occur in every phase of the study to strengthen external validity and confidently measure the effects of the intervention in a different context (Horner et al., 2005). Fidelity is needed to assess the consistency of intervention implementation and how this might affect the overall results (Horner et al., 2005; Wolery, 2013). If the intervention is not implemented with fidelity, this can create a weaker foundation for the functional relation between the intervention and the target behaviors (Wolery, 2013). Social validity is needed to assess the social reasons behind implementing a specific intervention as well as the stakeholders' opinions about treatment acceptable, efficiency, effectiveness, and continuation of the intervention outcomes when the study is completed (Horner et al., 2005; Reichow et al., 2008; Wolf, 1979). Without social validity measures, the entire study comes into question regarding the overall beneficial nature of this intervention and target behaviors for the participant and all other stakeholders (Horner et al., 2005; Reichow et al., 2008).

Each study's experiment(s) were analyzed using the evidence standards provided by the *WWC* (Kratochwill et al., 2010, 2013). This is necessary when considering the visual components of the data and how this plays into the overall effects. Visual analysis is commonly used to analyze the effects of single-case research and is recommended in addition to using effect sizes (Kratochwill et al., 2013). All studies except one either

partially met the standards (*Moderate Evidence)* or met all the standards (*Strong Evidence)*. This is encouraging when analyzing the overall effectiveness of interventions to promote employment skills because a majority of the studies that met (with or without reservations) the basic design standards also met or partially met the evidence standards. Even though some experiments received low scores for immediacy of change in level or baseline trend, the quality of single-case designs consistently revealed obvious positive treatment effects across a majority of experiments and studies. Overall positive effects could be the result of publication bias, or only publishing studies that show significant or visually unambiguous effects, but the overall search for articles in this review included both peer-reviewed and non-peer-reviewed articles to combat this bias (Cooper, 2010). Regardless, the bias may still exist, but with the present information, 38 studies and 75 experiments were found to have moderate or strong evidence for overall positive treatment effects.

These 38 studies were categorized into groups according to the type of intervention implemented. The interventions most commonly implemented were video modeling and visuals. Only video modeling interventions met the 5-3-20 evidence-based practice standards (Horner et al. 2005; Kratochwill et al., 2010, 2013). However, there are some limitations regarding the definitions of this intervention. Video modeling is a broad definition of this type of intervention because there are many types of models (e.g., peers, adults, self, or point-of-view, in-vivo) as well as different implementations (e.g., video priming vs. video prompting) that make it difficult to specifically determine the most effective component or variety of video modeling (Bellini & Akullian, 2007;

Mason et al., 2012a, 2013a). There were not enough studies included in our meta-analysis for a component analysis for video modeling to specify which type of component brought the strongest effects (Boles, 2015e). Therefore, the results of this evidence-based intervention should be analyzed with caution and future research should focus on studies analyzing the effects of specific components in these interventions.

When analyzing all studies that employed one of the four interventions, there was much author or institution overlap, which made it difficult to reliably analyze similar intervention effects in different contexts. Each of the interventions that did not meet evidence-base standards (i.e., audio cueing, visuals, and prompting) should be replicated across multiple authors and institutions to ensure intervention effectiveness and social validity across contexts as well as enhance the evidence base.

This review implemented a thorough quality analysis for all of the included studies, but there were limitations. First, all data collection was scored using rating scales, which made it difficult in some cases to have a high percent of IRR agreement. Although most of the rating scales only had 2 to 3 items, attaining reliability across raters was difficult with more abstract measures such as the descriptive design indicators. Difficulties arose between the raters because there were either too many components included in each item for the specified study element or each study reported the elements in different ways (e.g., used different jargon, reported it with different measures, reported in a different section of the paper where it was harder to find) which made it hard to discern the correct score for that study element.

Second, the lack of a sufficient number of studies that implemented audio cueing and prompting interventions made it difficult to make assumptions based on the quality and evidence of the intervention type. In addition, 7 out of the 38 studies did not fall into any intervention categories because the interventions were implemented as packages and included a wide range of intervention components. These intervention packages make it difficult to analyze the quality and efficacy of specific intervention components. Additionally, many of these studies had overlapping authors, which excludes the studies from being counted separately as additions to the evidence base for employment skill interventions.

Third, a broad spectrum of participants were included in these studies. The most common type of DD included ASD, ID, comorbidity of DD with other disabilities (e.g., ASD with ID), and multiple disabilities (i.e., included more than 2 diagnoses). Quality, effective, and evidence-based video modeling interventions included participants with the most diverse disabilities (i.e., ASD, ASD with ID, ID) and ages (i.e., 12 to 15 years old, 16 to 21 years old, and 22 years old and older). The diversity of participants in video modeling interventions made it difficult to determine this as an evidence-based practice for a specific population. Continuous research is needed regarding larger samples of participants with ASD, ASD with ID, and ID ages 12 year old and older to specify video modeling efficacy for certain individuals. Research is also needed for visual interventions regarding larger samples of participants with ASD and ASD with ID ages 12 to 15 years old and 22 years old and older.

This quality review added to the employment skill intervention literature base, but future research needs to focus on filling the existing gaps in this body of research. Researchers need to concentrate on replicating existing and promising interventions across institutions and authors for this adolescent and adult population with DD to increase the evidence base. More SCED research is needed for audio cueing/coaching and specified prompting interventions for individuals diagnosed with ASD, ASD with ID, and ID for future assessment of evidence-based practices regarding employment skills. Further, researchers need to thoroughly describe all participants, settings, implementers, procedures, and target behaviors; and include maintenance, generalization, fidelity and social validity measures according to the quality standards (CEC 2014; Horner et al., 2005; Kratochwill et al., 2013; Maggin et al., 2013; Reichow et al., 2008; and Wolery, 2013).

CHAPTER IV

THE EFFECTS OF POINT OF VIEW VIDEO MODELING IN TEACHING CONVERSATIONAL SKILLS TO HIGH SCHOOL STUDENTS WITH AUTISM SPECTRUM DISORDER

Autism spectrum disorders (ASD) are typified by deficits in communication, poor social interaction, and rigid thinking and behavior (Matson & Rivet, 2008). Deficits in social interaction may include inappropriate or lack of eye contact, inappropriate conversational turn-taking, poor or no reciprocal communication, or disinterest in the conversational partner (Adreon, 2007). These impairments lead to challenges initiating conversations, engaging in reciprocal exchanges, and inferring the feelings and interests of others. This can be particularly detrimental to adolescents with ASD as social competence is essential to navigating less structured environments and developing social connections (Allen, Wallace, Renes, & Bowen, 2010b; Glennon, 2001). In addition to the social complexities of adolescents, cognizance of one's idiosyncrasies can lead to further withdrawal resulting in the development of comorbid anxiety and depression (Mazurek & Kanne, 2010; Ratto, Turner-Brown, Rupp, Mesibov, & Penn, 2010; Wood et al., 2009). Given the negative impact of impairments in socio-communicative skills, identification of effective interventions is necessary to increase successful peer engagement and improve quality of life.

Video-based modeling is one such intervention with empirical support for improving targeted skills for individuals with ASD (Bellini & Akullian, 2007; Mason et

al., 2012a; Mason, Ganz, Parker, Boles, Davis, & Rispoli, 2013b; Shukla-Mehta, Miller, & Callahan, 2010; Van Laarhoven, Zurita, Johnson, Grider, & Grider, 2009). However, most research has focused on implementing video modeling interventions to improve independent living skills or play skills (McCoy & Hermansen, 2007). Little research has explored the use of this intervention in teaching conversational or other socio-communicative skills (e.g. eye contact and body orientation). Video modeling requires creating a video of an accurate exemplar of the target behavior. The video can be created by either coaching the individual through the task/behavior and editing out the prompts, known as video self-modeling, or video recording another individual acting out the desired skill, known as video modeling with other (e.g. peer, sibling, adult) as model (Bellini & Akullian, 2010; Mason et al., 2012a). Another form of video modeling, point-of-view video modeling (POV), may be used to minimize distracting stimuli often present in video self-modeling or video modeling other (Tetrault & Lerman, 2010). POV is defined as filming a completion of a specified target skill from the perspective of the individual viewing the video. The individual modeling the behavior is not shown except for the needed tools for completing the task (e.g., hands) or nothing at all if the viewpoint is strictly within eye level of the perspective of the model (Hine & Wolery, 2006; Tetrault & Lerman, 2010).

Implementation of POV has advantages over the frequently utilized third-person perspective in the video-based modeling literature. Due to the nature of POV, the video exemplar focuses only on the task being completed, as the video is recorded exclusively from the view of the individual completing the activity (Tetrault & Lerman, 2010). This

alone increases the ease of preparation, in terms of creating the video, as the actor will likely require less preparation or training in comparison to video modeling other, and editing will be minimized due to the clarity of the target skill being modeled in comparison to video self-modeling (Shipley-Benamou et al., 2002; Rai, 2008). Additionally, it is possible for the one creating the video to also serve as the model, through the use of headgear and/or earpiece video camera, reducing the personnel resources required for production. Also, filming from a first-person perspective omits unnecessary stimuli that may inadvertently detract from the viewer's focus on the targeted skill (Rayner et al., 2009).

Typically video-based modeling interventions, including POV, are implemented as part of a package rather than alone, making it difficult to ascertain to which intervention component (e.g., video-based modeling, prompting, reinforcement) to attribute improvements in targeted skills (Shukla-Mehta et al., 2010). In a meta-analysis of single-case research implementing POV, Mason et al. (2013a) found differential effects when POV was delivered alone, with reinforcement, or as part of a package. Although POV alone was found to have statistically significant stronger effects, only three studies (Cannella-Malone et al., 2006; Cannella-Malone et al., 2011; Hine & Wolery, 2006) included in the meta-analysis implemented POV alone, thus limiting the confidence in these results. Two of those studies (Cannella-Malone et al., 2006; Cannella-Malone et al., 2011) compared the efficacy of POV alone with POV as part of a prompting procedure including error correction to improve independent living skills for adolescents with ASD. Although the POV alone procedure did result in an increase

in the number of steps completed accurately for setting the table or buying groceries, the POV package with prompting and error correction resulted in a higher percentage of steps completed accurately. Hine and Wolery (2006) implemented POV alone to teach imitative play skills to two children, ages 30-43 months, with ASD. The POV alone intervention yielded an increase in the number of modeled steps each child imitated for gardening play; however, one participant required the addition of a verbal prompt to imitate the video and reinforcement contingent on imitation of modeled steps for cooking play. More research, particularly for other targeted skills, such as socio-communicative skills, is necessary to evaluate the efficacy of POV alone.

POV has primarily been implemented with individuals with ASD to teach independent living skills (Rayner et al., 2009) such as cooking related tasks (Sigafoos et al., 2005; Shipley-Benamou et al., 2002) dressing (Norman, Collins, & Schuster, 2001), and household tasks (Ayres & Langone, 2005; Shipley-Benamou et al., 2002). Additionally, POV has been implemented to improve play skills (Hine & Wolery, 2006), decrease tantrumming behavior (Schreibman, Whalen, & Stahmer, 2000), and improve successful transitions (Vannest, Parker, & Gonen, 2011).

Only one study has explored the functional relationship between POV and improvements in socio-communicative skills for individuals with ASD (Tetrault & Lerman, 2010). Specifically, Tetreault and Lerman (2010) utilized POV to increase social initiations and eye contact for 3 children, ages 4-8, with ASD. In addition to POV, reinforcement was delivered contingent on attention to the video and successful implementation of targeted skills. The intervention did result in increases in targeted

skills for two of the participants. Increases were observed for the third participant when a least-to-most prompting procedure was added to the intervention package. There are no published studies that explore the use of POV to increase socio-communicative skills for adolescents with ASD. Given the simplicity of the intervention, POV appears to be a promising avenue for improving socio-communicative skills for adolescents with ASD; however, evaluation of the efficacy of POV for improving socio-communication for adolescents with ASD is an area in need of more scientific evidence.

Given the limited evidence regarding the use of POV implemented alone rather than as part of a package as well as the lack of empirical evidence supporting the use of POV to increase socio-communicative skills for adolescents with ASD, more research is needed. This study seeks to address both of these limitations by answering the following research question: Is there a functional relationship between the implementation of POV alone and improvements in socio-communicative skills for adolescents with ASD?

Methods

Participants

Two male high school students initially participated in the study. These students were attending a social skills class to assist them with socio-communicative skill deficits for 55 minutes every day. The teacher of this class reported on the current skill level and diagnosis of each participant. John was 19 years old at the beginning of the study, and he was diagnosed with high functioning autism. John worked at a local fast food restaurant. According to his teacher's report, John could hold a conversation with others, but he only talked about topics he was interested in and did not comment on statements made by others. John interrupted others if the conversation was not about his interests, or if the

person was involved in another conversation. John did occasionally ask a question of the other individual in the conversation, but the question usually pertained to the subject that he was most interested in. When interacting with others, he used minimal appropriate eye contact.

Steven was 17 years old at the time of the study. He was diagnosed with high functioning autism. His teacher reported that he exhibited challenging behaviors when he felt annoyed with others by yelling at them or getting in their personal space. When having conversations with others, the conversation usually focused on his specific interests. If the subject being discussed was not interesting to him, he would interrupt the other individual and keep talking about his own interests. Steven would vary rarely make eye contact with others or ask questions regarding the other person's interests.

Setting and Materials

All sessions took place in a social skills classroom arranged similarly to a typical high school classroom with medium-sized tables with two chairs to each table. This 55-minute class was considered an elective course, and in addition, the participants attended general core courses for the remainder of the school day. Sessions were conducted in a quiet corner in the back of the classroom during small group and individual instruction. The classroom contained nine tables with two chairs to each table, the teacher's desk at the front left corner, and a projector with a pull down screen in the middle of the front of the room. There were six students attending the social skills class, and one teacher taught the class. Five of the students were diagnosed with ASD and one student was diagnosed with an intellectual disability. The lessons taught focused on independent living skills

(e.g. budgeting finances and hygiene), socio-communicative skills (e.g. how to compliment someone), and job skills (e.g. interviewing, what to wear to an interview, how to fill out a job application). The target skills of the previous and current class lessons did not overlap with the target skills focused on in this video modeling intervention.

Video models. A handheld video camera was used when filming the POV clips. A research assistant filmed each conversational skill video using the camera to record his perspective, like that of the participants', as he sat down to have a conversation with a conversation partner. Each POV clip used this filming technique to show the participant performing the skill from his perspective. Due to the complexity of the target socio-communicative skills, arrows, text, or symbols were added in each video clip to direct the participant's attention to the specific skill being modeled. Each video included the conversation partner (the first author in this case) that was typically present during the baseline and intervention study sessions. The conversation partner did not prompt the participant to perform target behaviors during the video clip. A hand-held device, the size of a smartphone, was used to present these video clips to the participants.

A video clip was created for each of the three target skills: (a) eye contact; (b) body orientation; and (c) withholding interruption. The eye contact video clip included a model, form the participant's point of view, of appropriate eye contact towards a conversation partner. The video clip began with the title "Eye Contact" appearing for 3 seconds, and then the conversation partner came into view. The conversation partner began asking questions such as, "how was your weekend?", or "how is school going?",

while the participant model answered the questions with phrases such as, “I met a friend after school and we hung out”. To represent appropriate eye contact during the conversation, the camera was focused on the conversation partner’s eyes (i.e. eye contact) for a 3 to 5 second interval, and then, the camera shifted to point away from the conversation partner’s eyes for an interval of about 3 seconds (i.e. break in eye contact). This appropriate break in eye contact is necessary to avoid “staring” at the conversation partner for a socially inappropriate length of time. The eye contact and break in eye contact intervals continued to alternate for the remainder of the video clip. A small green arrow was included to provide the participant with the suggested point of eye contact when speaking with a conversation partner. The green arrow radiated from the perspective of the participant and pointed in the direction that the participant should focus his eyes. The eye contact video clip lasted a total of 1 minute and 9 seconds.

The body orientation video clip included a model, from the participant’s point of view, of appropriate body positioning when sitting across from a conversational partner. The video clip began by presenting the title, “Body Position Towards Another Person”, and then the camera began to move, representing the participant walking and sitting down opposite the conversation partner. The video clip included arrows that radiated from the participant’s head, shoulders, and knees to indicate that these body parts should face toward the conversation partner while in a sitting position. The words “head”, “shoulders”, and “knees” were written above each arrow. The body orientation video clip lasted a total of 15 seconds.

The withholding interruption clip included a model, from the participant's point of view of inappropriate and appropriate examples of contributing to a conversation being held by two other conversation partners. The video presented the title "Interrupting Others" and then the two conversation partners appeared. The two conversation partners conversed about the day by asking each other "how was your day today?", and then answering each other with phrases such as, "it was good, but I was really busy". Towards the middle of this conversation, the participant model interrupted (i.e. began to speak while another individual was currently speaking) the conversation between the two other individuals. With this interruption, a 'no' symbol (red circle with a diagonal line through it) popped up on the screen. The second part of the clip consisted of the participant model waiting until both conversation partners were silent to make a statement relative to the topic being discussed. After this appropriate contribution to the conversation, a green checkmark symbol appeared. The withholding interruption video clip lasted a total of 30 seconds.

Experimental Design and Measurement

The investigators implemented a multiple-baseline, single-case research design across conversational skills for each participant. The study was carried out over the course of approximately 8 weeks with 2 to 4, 2-min sessions per week.

Measurement. Live data was collected during every session. Sessions occurred one to two days a week, with exceptions for school events or holidays, for over three months. The frequency of data collection varied for every skill based on the occurrence

of variability or stability in baseline and intervention phases. Stable data lead to less frequent probes of each skill throughout the current phase.

Dependent variable. Target skills were measured utilizing rating scales developed by the first author. The rating scale for each conversational skill was developed based on observations of the skills presented during conversations between typically developing individuals (see Table E1 in Appendix E). Mason, Rispoli, Ganz, Boles, and Orr (2012b) used rating scales to measure socio-communicative skills such as eye contact, turn taking, and sharing emotions. These rating scales were utilized as a guide for the development of the rating scales for use in this study. A rating scale is important to measure socio-communicative skills due to complexity of these skills (i.e. multiple components). Frequency data or interval recording is not ideal for data collection due to the number of complex elements (e.g. eye contact with the conversation partner, looking away and not staring, giving eye contact to the conversation partner again after looking away quickly) that make up the skill. The rating scale for eye contact, which was modified based on the scale developed by Mason and colleagues (2012a), and the scale for avoiding interruption utilized a 5-point Likert-like scale. The scale for body orientation utilized a 3-point Likert-like scale. The decision to utilize a 3-point scale as opposed to the 5-point scale utilized for the other two behaviors was due to the fact that body orientation was defined as lining up three body parts (i.e. head, shoulders, knees) with the conversation partner. Therefore, only three ratings were required to measure the total number (i.e. 1, 2, or 3) of body parts lined up with the conversation partner. Table E1 provides a list of the target behaviors with anchors for each scale clearly defined.

Social Validity. Two surveys were distributed to the participants during the course of the study. The first author created these two surveys to gauge student social awareness and the social validity of the study. The pre-baseline survey was given to the students before baseline sessions began and contained eight questions pertaining to feeling comfortable or nervous in specific social situations (see Table E2). Both John and Steven revealed that it was somewhat difficult to talk to others, and they felt nervous sometimes in social situations. They also indicated that the way they felt towards others in conversation depended on how well they knew the person and "what kind of person" he or she was. In summary, both John and Steven had only minor concerns regarding social situations. The second survey was given at the end of intervention sessions and contained six questions to document the social validity of the POV intervention (see Table E3). The results of this survey will be discussed further in the *Social Validity* subsection of the *Results* section.

Inter-observer agreement. For interrater reliability, the first author and other graduate students in the same special education doctoral program, with previous research experience and published works, independently collected data immediately following each session. The first author explained and modeled the data collection procedures for all graduate students participating in data collection. Additional raters did not interact in the conversation unless addressed by the participant. Also, any additional rater sat in a chair that was a few feet to the side of the first author who primarily interacted in the conversations. All data collectors participated in previous observations and 'get to know you' sessions with the participants prior to the start of the study in order to avoid novelty

at the start of baseline sessions. Participants were allowed to talk to either data collector during the conversation, but the first author asked the scripted questions during all conversations. These procedures were put in place to facilitate more natural conversations and create a comfortable environment for the participants with additional raters present.

Interrater reliability (IRR) for John's data was taken for a total of 54% of eye contact, 62% of body orientation, and 90% of interrupting data sessions across baseline and intervention phases. During baseline sessions, IRR was taken for 20%, 43%, and 83% of the sessions for respective target skills. During intervention sessions, IRR was taken for 71%, 83%, and 100% of the sessions for respective target skills. IRR for Steven's data was taken for a total 36% of eye contact, 38% for body orientation, and 55% of sessions for interruption data sessions across baseline and intervention phases. During baseline sessions, IRR was taken for 25%, 14%, and 80% of sessions for the respective target skills. During intervention sessions, IRR was taken for 43%, 67%, and 25% of sessions for the respective target skills. IRR was calculated as percentage of observed agreement while giving half credit for near misses between the two raters. Near misses are defined as the second raters scoring the skill with only one rating (0.50 for eye contact and interruption, and 1.00 for body orientation ratings) higher or lower than the first observer's given rating for that skill during a session. Percentage observer agreement with credit to near misses was calculated using the PABAK-OS (Vannest et al., 2011). John's IRR for each target behavior was 75%, 81%, and 94% for eye contact, body orientation, and interrupting measures respectively. Overall, John's IRR measured

83% agreement. Steven's IRR were calculated as 100%, 100%, and 90% for each measure, respectively. Overall, Steven's IRR measured 97% agreement.

***Treatment Integrity*.** Procedural integrity was recorded as a yes or no correct occurrence of each step in the intervention process defined below in the *Procedures* section. A graduate student observed the first author and scored implementation of the intervention. John's procedural integrity was collected for 43% of eye contact intervention sessions, 50% of body orientation intervention sessions, and 100% of interruption intervention sessions. Steven's procedural integrity was collected for 29%, 50%, and 20% during the respective conversational skill intervention sessions. There was 100% procedural integrity across each conversational skill intervention for each participant. This high percentage of procedural integrity may be attributed to the scripted nature of the intervention.

Procedures

Pre-baseline sessions. Three short preliminary sessions were conducted with each participant before baseline and consisted of talking with participants about their interests and simultaneously, informally analyzing social conversational skill strengths and deficits. These introductions and conversations with the participants were also conducted to control for reactivity due to novel researchers in the classroom. Participants were made to feel comfortable with each researcher before any baseline or intervention sessions were conducted. For all sessions after pre-baseline, only one researcher served as the conversation partner for the participants in eye contact and body orientation sessions. Two researchers were required to be present during the interruption sessions to

create the scenario of a conversation between the researchers that would allow for interruption or the withholding of interruption by the participant.

Baseline. During baseline sessions one or two researchers asked one of the participants to join them in the back corner of the room to take part in a conversation. The researcher asked the participant what he would like to talk about and narrowed it down to discussing his weekend, schoolwork, interests, or how his school day had gone so far. Each session lasted 2 minutes. Each session began with one researcher saying, "tell me about your day/weekend/interests". The first author had a stopwatch that marked the end of data collection after 2 minutes without interrupting the participant if he was still having a conversation with the researcher. The stopwatch was started when the participant started talking to the researcher after he was asked about his day/weekend/interests. When the conversation was over, the participant was thanked \for talking to the researcher and asked to resume participating in the regularly scheduled class activity. Introduction of the POV model for the first skill, eye contact, occurred once baseline data indicated a clear and stable pattern of behavior.

POV intervention. The POV model for each skill was systematically introduced at 3 different points in time for each participant. Proceeding skills were introduced once data indicated a predetermined criterion for improvement from baseline and stability. Improvement from baseline criterion was predefined as three concurrent sessions with a minimum half-point increase in rating from the mean score in the baseline phase (see Table E1 for measurement scale) for eye contact and interruption and 1-point for body orientation. Once the participants met criterion for each conversation skill, the next

conversation skill POV intervention was implemented while the intervention continued for the previous skill.

The POV intervention sessions were conducted similarly to baseline sessions. However, before the conversation began, the first author told the student, “we are going to watch a video”. The participant was then told, “this video is shown from your perspective, do you know what that means?”, and if he said that he did understand, the researcher moved on. If the participant said that he did not understand, the researcher told the participant, “it means that the way the camera moves is the way that your eyes would move”, and “the person in the video is the person you would be talking to and looking at”. Then, the researcher pushed play and handed the smartphone to the participant. After the video clip was completed, the device was put out of sight and the researcher asked the participant, “tell me about your day/weekend/interests”. The videos were played before each session relating to the target behavior, which is considered as video priming. Procedural fidelity was taken for each of these steps by another researcher, and it was scored as an occurrence or non-occurrence of the correct follow-through of each step.

Modifications. Steven required a slight modification solely for the withholding interruption skill intervention phase. At the beginning of the intervention phase, he did not show any improvement in withholding interruption. During the third session, a modification was introduced that included a decrease in the time limit of the conversation sessions to 1 minute. As a result, a small increase in skill rating was achieved in the third session, but it was not sustained. During the fourth intervention

session, Steven began interrupting multiple times. No further modifications were made due to the ending of the school year, which created limited data collection opportunities.

Analysis

Data were analyzed using visual analysis consisting of variability, level, and trend of the data across baseline and intervention phases (Kazdin, 2011). Each conversation skill was analyzed individually as well as in tandem with the other conversational skills included in the multiple-baseline design.

Statistical analyzes included the use of Tau-U, a nonparametric measure, as an effect size measurement for each conversational skill. Tau-U is defined as a non-overlap effect size analysis that can control for trend in the baseline and intervention phases (Parker et al., 2011b). For this study, Tau-U was calculated as non-overlap for all six baseline and intervention phase contrasts. Tau-U scores range from -1.00 to 1.00 with scores above 0.00 indicating improvement from baseline to intervention phases. These scores were calculated using an online Tau-U calculator (Vannest et al., 2011).

Results

Figures E1 and E2 in Appendix E illustrate the results for each of the three conversational skills targeted for John and Steven. John's overall results show higher ratings in intervention phases compared to the ratings seen in baseline phases across all three skills. Steven's overall results indicate higher skill ratings in intervention phases in comparison to baseline phases for two conversational skills. The overall results indicate a moderate functional relation due to the absence of Steven withholding interruption during the intervention phase.

John's eye contact data indicated higher variability and slight increase in trend in the baseline phase compared to stability in the intervention phase with an almost flat trend. There was an immediate improvement in consistency and quality of eye contact from a rating of 2.50 in the last baseline session to a rating of 3.00 in the first session of intervention, which indicated a clear change in level from baseline (*M*=2.25) to intervention (*M*=2.93). Tau-U was calculated for the eye contact phase contrast and resulted in a strong effect size of 0.93. For the second target skill, body orientation, data indicated less variability in the baseline data with a slight decrease in trend in comparison to stability and a flat trend in the intervention phase. There was an immediate increase in body orientation towards the conversation partner from the last session of the baseline phase to the first session of the intervention phase, indicating clear change in level from baseline (*M*=2.00) to intervention (*M*=3.00). The body orientation phase contrasts resulted in a strong Tau-U effect size of 0.86. In the third target skill, interruption, data in the baseline phase were less variable with a slightly decreasing trend, and the data in the intervention phase were stable with a flat trend. There was a high rate of interrupting during baseline conversations (*M*=1.25); however, this improved to no evidence of interrupting in intervention conversations (*M*=3.00), indicating a clear change in level. The interrupting phase contrasts resulted in a strong Tau-U effect size of 1.00, indicating no overlap between baseline and intervention phases.

Steven's eye contact data indicated stable ratings of 1.00 with flat trend for all sessions in baseline (*M*=1.00). Intervention data were less variable and became more

stable starting in the fourth session with a slightly decreasing trend. An immediate improvement in consistency and quality of eye contact occurred from a rating of 1.00 in the last session of baseline to a rating of 2.50 for the first session in the intervention phase, which indicated a clear change in level from baseline (*M*=1.00) to intervention (*M*=2.14). The eye contact phase contrasts revealed a strong Tau-U effect size of 1.00. Body orientation data indicated a slight increase in trend and less variability in baseline with only one session with a rating of 1.00, a lack of any body orientation, out of seven total sessions, six of which had a rating of 2.00. Intervention data indicated a flat trend with stable ratings of 3.00 for all sessions. There was immediate improvement and a clear change in level in body orientation towards the conversation partner from baseline (*M*=1.90) to intervention (*M*=3.00). The body orientation phase contrast resulted in a strong Tau-U effect size of 1.00. Interrupting data indicated stability and a flat trend for baseline data at a rating of 1.00, indicating high rates of interrupting. Intervention data indicated less variability and a slightly increasing trend from one session out of four with a rating of 1.50, revealing a slight decrease in the rate of interrupting for that single session. The change in level was minimal between baseline (*M*=1.00) and intervention (*M*=1.13). The results in the interruption data indicated little change in ratings between the baseline and intervention phases. The interruption phase contrasts resulted in a weak Tau-U effect size of 0.25, which indicated very little improvement between baseline and intervention phases.

Social Validity

Social validity was measured via a survey asking questions about the participants' reactions to the video modeling intervention used in this study (see table E3). Both participants reacted positively to using portable technological devices to watch the videos, but they only wanted some of the videos put on their personal devices. Both participants also stated that they would use these same videos to watch before they stepped into certain social situations. However, the participants did mention their dislike for watching the same video over again for back-to-back sessions, and Steven commented that some of the videos were 'common sense'.

Discussion

Overall, John improved in all three conversational skills resulting in strong Tau-U effects sizes while Steven improved in two out of the three conversational skills with strong Tau-U scores as well. Both participants were satisfied with the intervention and showed interest in using this POV intervention during their typical daily routines.

John immediately improved in all conversational skills during intervention phases in comparison to baseline measures. The increases in skill ratings remained high during all intervention phases. Steven's skill ratings immediately increased for eye contact and body orientation when compared to baseline measures. These higher skill ratings remained constant throughout the intervention phases with the exception of the interruption conversational skill. The overall functional relation between POV (alone) and the targeted intervention skills yielded strong effects for eye contact and body

orientation skills for both participants, a strong effect for interruption skills for John, and a weak effect for the interruption phase for Steven.

Steven continued to interrupt multiple times in the two-minute period even after the video modeling intervention. This occurrence of behavior may have been due to the 2-minute length of the time period that Steven would have been prompted by the video modeling clip to withhold from interrupting the conversation before being able to participate in the conversation. The time limit may have needed to be decreased to an individualized level for Steven due to his urgency to share his perseverative interests with others. This situation supports the importance of creating individualized intervention plans to promote success in teaching individuals specific skills. Another possible reason for Steven's continued interrupting behaviors may be due to the complexity of the interruption video model containing an example of the participant interrupting and then of the participant not interrupting the conversation being held by the researchers. This may have created confusion or difficulty in the expected behaviors that were being modeled from the participant's point of view.

These results were similar to the increases in social initiations and eye contact seen in Tetreault and Lerman (2010) as a result of the implementation of POV with prompting and reinforcement procedures for younger children. The current study is unique in that POV was implemented for older individuals without the use of prompting or reinforcement procedures. Therefore, the results show a promising future for the use of POV for older individuals with ASD.

These results may be due to some factors that were not controlled for in this study. Positive results may have been achieved because of the functioning levels of the individuals. Both individuals were diagnosed with high functioning autism and had some socio-communicative skill training in previous social skills classes. This may have aided the participants in attending to and quickly reacting to the video models positively. Both participants were very interested in technology, which could be a reason that they reacted so positively to using a smartphone as a medium for the video models.

The social validity measures for this study yielded promising results. When given the pre-study survey, both participants mentioned nervousness when talking to others as well as difficulty making eye contact. This intervention aided the participants in expressing these target behaviors in social situations that typically brought anxiety. Overall, these participants reported a desire to use these video models in everyday activities. This information represents the acceptability and preference of this intervention for these participants and it presents a possibility of this intervention being used as a tool to promote generalization of target skills.

There were some limitations present in the current study. First, participant information was limited. Additional information (e.g., adaptive assessment scores, source of diagnosis, additional special education services provided) would have been beneficial for future replication of this intervention with similar participants. Additionally, IRR across target behaviors was low at 75% and 81% for eye contact and body orientation respectively. This could be due to the general complex nature of measuring socio-communicative skills with a rating system. Another limitation was due

to the schedule of the school regarding testing and dismissal for the summer. The researchers were limited in the amount of data and number of sessions that could be implemented. Given more time, the researchers could have continued Steven's intervention phase data for the interruption conversation skill to determine if a lengthier intervention would have had better results. Further, the time limit did not allow maintenance data collection for the targeted conversational skills for both individuals. Another limitation was the absence of generalization measures. All data were collected in the classroom with the same conversation partner. Generalization measures would have aided the researchers in assessing the usability of this intervention in different settings and in the presence of different conversational partners.

Future research with the understudied population of high-functioning adolescents and young adults with ASD should include investigations of POV applied to a wide range of skills tailored to individuals with ASD and other developmental disabilities. Future research should focus on maintenance and generalization measures to enhance the use of POV for long-term purposes and in multiple contexts. Some important skills to assess in the future may include other social, independent living, and career skills. It may also be beneficial to compare the use of different technological mediums to present the video models due to ever-changing technology. With the growing research in video modeling in general, it may be beneficial to compare the effectiveness of POV (alone) with other forms of video modeling, such as, self-modeling and adult or peer modeling while measuring which interventions may be more acceptable and practical in the presented context.

CHAPTER V

CONCLUSION

Many individuals diagnosed with developmental disabilities (DD) have difficulties in acquiring or maintaining employment after high school due to employment skill deficits (Carter et al., 2012; Grigal & Deschamps, 2012; Hanley-Mazwell & Izzo, 2012; Hendricks, 2010; Hendricks & Wehman, 2009; Wehmeyer, 1994). Compared to their typically developing peers, less than 50% of individuals with DD are competitively employed or maintain employment at their current job (Newman et al., 2009, 2011; Sanford, Newman, Wagner, Cametoo, & Knokey, 2011). For many of these individuals, unemployment can lead to increased dependence on others, insufficient income, or lower quality of life (Billstedt et al., 2005; Hughes, 2001; Wehmeyer, 1994).

To expand the literature base of employment skill interventions for individuals with DD, efficacy and quality analyses were conducted in three separate articles in this dissertation. The meta-analysis (first article) addressed research gaps in employment-related literature by analyzing effect sizes based on moderator variables to inform researchers and practitioners of employment skill interventions' effectiveness with a variety of components. The quality analysis (second article) put all studies that implemented employment skill interventions under methodological scrutiny to inform researchers of the need to raise the quality standards for future research and to inform practitioners of current quality and evidence-based studies. The single-case study (third article) implemented a point-of-view video modeling intervention to teach individuals

with ASD social skills that can be used in a variety of contexts (e.g., employment, home, school). This study contributed to the literature of point-of-view video modeling interventions for understudied high school age populations of individuals with ASD.

The results of each study in this dissertation provide a foundation for the progression of future research and practice. There is a need for more research in effective and evidence-based interventions to promote employment skill acquisition and performance for understudied populations of adolescent and adult individuals with DD. The outcomes in this dissertation indicate the need for continued research in the efficacy of video modeling, audio cueing, visual, and prompting employment skill interventions as well as the impact of dependent variables, participant characteristics, settings, and interventionist characteristics as moderator variables. Future research should also focus on video modeling intervention components (i.e., video prompting, video priming, self-modeling, peer modeling, adult modeling, point-of-view modeling, and in-vivo modeling) in relation to the moderate to strong effects yielded for individuals with DD. Researchers need to thoroughly assess the contributing factors to current employment-related intervention efficacy and the benefits of these interventions in relation to helping individuals acquire and maintain jobs.

In practice, the results from all three articles allow for teachers, parents, employers, job coaches, and other stakeholders to make more informed decisions based on intervention efficacy, quality and evidence-base. Video modeling interventions were seen to be effective across a majority of target behaviors, task complexities, participant age groups, participant IQ levels, participant developmental disability diagnoses,

intervention settings, and interventionist characteristics. As video modeling is a mobile intervention that can be practically applied by the individual with a disability, it is promising that this intervention can be effectively used in the work place to promote independence and maintenance of acquiring and performing employment skills (Bellini & Akullian, 2007; Mason et al., 2012a, 2013a). All outcomes provide pertinent information regarding job acquisition and maintenance for all stakeholders (e.g., teachers, individuals with DD, job coaches, employers) to decrease rates of unemployment for individuals with DD (Newman et al., 2009, 2011; Sanford et al., 2011).

Overall, the results of the analyses in this dissertation provided evidence for effective, quality, and evidence-based interventions to teach individuals with DD a variety of employment skills. This is promising when addressing issues surrounding the transition processes in special education. A diverse group of participants, study elements, and behaviors can provide a more solid foundation for the individualization and effectiveness of employment skill interventions for future research and practice.

REFERENCES

Adreon D. (2007). Evaluating the college transition needs of individuals with high-functioning autism spectrum disorders. *Intervention in School and Clinic,42*, 271–279.

Allen, K.D., Wallace, D.P., Greene, D.J., Bowen, S.L., & Burke, R.V. (2010a). Community-based vocational instruction using videotaped modeling for young adults with autism spectrum disorders performing in air-inflated mascots. *Focus on Autism and Other Developmental Disabilities, 25*(3), 186-192. doi: 10.1177/1088357610377318

Allen, K.D., Wallace, D.P, Renes, D., Bowen, S.L., & Burke, R.V. (2010b). Use of video modeling to teach vocational skills to adolescents and young adults with autism spectrum disorders. *Education and Treatment of Children, 33*(3), 339-349.

Allen, K.D., Burke, R.V., Howard, M.R., Wallace, D.P., & Bowen, S.L. (2012). Use of audio cuing to expand employment opportunities for adolescents with autism spectrum disorders and intellectual disabilities. *Journal of Autism and Developmental Disorders, 42*, 2410-2419. doi: 10.1007/s10803-012-1519-7

Alwell, M., & Cobb, B. (2009). Functional life skills curricular interventions for youth with disabilities: systematic review. *Career Development for Exceptional Individuals*, 32, 82-93. doi: 10.1177/0885728809336656

Ayres, K. M., & Langone, J. (2005). Intervention and instruction with video for students with autism: A review of the literature. *Education and Training in Developmental Disabilities*, *40*, 183–196.

Banda, D.R., Dogoe, M.S., Matuszny, R.M. (2011). Review of video prompting studies with persons with developmental disabilities. *Education and Training in Autism and Developmental Disabilities, 46*(4), 514-527.

Bellini, S., & Akullian, J. (2007). A meta-anlaysis of video self-modeling interventions for children and adolescents with autism spectrum disorders. *Exceptional Children, 73,* 264-287.

Bennett, K., Brady, M.P, Scott, J., Dukes, C., & Frain, M. (2010). The effects of covert audio coaching on the job performance of supported employees. *Focus on Autism and Other Developmental Disabilities, 25*(3), 173-185. doi: 10.11771088357610371636

Bennett, K.D., Rangasamy, R., & Honsberger, T. (2013a). The effects of covert audio coaching on teaching clerical skills to adolescents with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 43*, 585-593. doi: 10.1007/s10803-012-1597

Bennett, K.D., Rangasamy, R., & Honsberger, T. (2013b). Further examination of covert audio coaching on improving employment skills among secondary students with autism. *Journal of Behavioral Education, 22*, 103-119. doi: 10.1007/s10864-013-9168-2

Bereznak, S., Ayres, K.M., Mechling, L.C., & Alexander, J.L. (2012). Video self-prompting and mobile technology to increase daily living and vocational independence for students with autism spectrum disorders. *Journal of Developmental and Physical Disabilities, 24*, 269-285. doi: 10.1007/s10882-012-9270-8

Billstedt, E., Gillberg, C., & Gillberg, C. (2005). Autism after adolescence: Population-based 13- to 22-year follow-up study of 120 individuals with autism diagnosed in childhood. *Journal of Autism and Developmental Disorders, 35*(3), 351-360. doi: 10.1007/s10803-005-3302-5

Boles, M. (2015a). Basic Quality Design Standards Rating Scale for Single-case Design. Available electronically from http://hdl.handle.net/1969.1/153769.

Boles, M. (2015b). Descriptive Quality Indicators Rating Scale for Single-case Design. Available electronically from http://hdl.handle.net/1969.1/153770.

Boles, M. (2015c). Additional Phase and Assessment Descriptive Quality Indicators for Single-case Design. Available electronically from http://hdl.handle.net/1969.1/153771

Boles, M. (2015d). Visual Analysis/Evidence of Effect Quality Indicators for Single-case Design. Available electronically from http://hdl.handle.net/1969.1/153772.

Boles, M. (2015e). *Effective Interventions In Teaching Employment Skills to Individuals with Developmental Disabilities: A Single-Case Meta-Analysis*. manuscript in preparation.

Bowman-Perrott, L., Davis, H., Vannest, K., Williams, L., Greenwood, C., & Parker, R. (2013). Academic benefits of peer tutoring: A meta-analytic review of single-case research. *School Psychology Review, 42*(1), 39-55.

Brossart, D.F., Vannest, K.J., Davis, J.L., & Patience, M.A. (2014). Incorporating nonoverlap indices with visual analysis for quantifying intervention effectiveness in single-case experimental designs. *Neuropsychological Rehabilitation, 24*, 464-491. doi: 10.1080/09602011.2013.868361

Camargo, Siglia (2012). *Behaviorally-Based Interventions for Improving Social Interaction Skills of Children with ASD in Inclusive Settings: A Systematic Review and Meta-Analysis.* Doctoral dissertation, Texas A&M University. Retrieved from http://hdl.handle.net /1969.1/148106.

Cannella-Malone, H. I., Fleming, C., Chung, Y. C., Wheeler, G. M., Basbagill, A. R., & Singh, A. H. (2011). Teaching daily living skills to seven individuals with severe intellectual disabilities: A comparison of video prompting to video modeling. *Journal of Positive Behavior Interventions, 13, 144-153.* doi:10.1177/1098300710366593

Cannella-Malone, H. I., Sigafoos, J., O'Reilly, M., De La Cruz, B., Edrisinha, C., & Lancioni, G. E. (2006). Comparing video prompting to video modeling for teaching daily living skills to six adults with developmental disabilities. *Education and Training in Developmental Disabilities*, *41*, 344–356.

Carter, S.L. (2010). *The social validity manual: A guide to subjective evaluation of behavior interventions.* San Diego, CA: Elsevier.

Carter, E.W., Austin, D., & Trainor, A.A. (2012). Predictors of post-school employment outcomes for young adults with severe disabilities. *Journal of Disability Policy Studies, 23*, 50-63.

Carter, E. W., & Wehby, J. H. (2003). Job performance of transition-age youth with emotional and behavioral disorders. *Exceptional Children, 69*, 449 – 465.

Cavkaytar, A. (2012). Teaching café waiter skills to adults with intellectual disability: A real setting study. *Education and Training in Autism and Developmental Disabilities, 47*, 426-437.

Chandler, W., Schuster, J.W., & Stevens, K.B. (1993). Teaching employment skills to adolescents with mild and moderate disabilities using constant time delay procedure. *Education and Training in Mental Retardation, 28*, 155-168.

Chang, Y-J., Kang, Y-S., & Huang, P-C. An augmented reality (AR)-based vocational task prompting system for people with cognitive impairments. *Research in Developmental Disabilities, 34*, 3049-3056.

Cihak, D.F., Alberto, P.A., Kessler, K.B., & Taber, T.A. (2004). An investigation of instructional scheduling arrangements for community-based instruction. *Research in Developmental Disabilities, 25*, 67-88.

Cihak, D.F., Kessler, K.B., & Alberto, P.A. (2007). Generalized use of a handheld prompting system. *Research in Developmental Disabilities, 28*, 397-408. doi: 10.1016/j.ridd.2006.05.003

Cihak, D.F., Kessler, K., & Alberto, P.A. (2008). Use of a handheld prompting system to transition independently through vocational tasks for students with moderate and

severe intellectual disabilities. *Education and Training in Developmental Disabilities, 43*(1), 102-110.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37-46. doi: 10.1177/001316446002000104

Connis, R.T. (1979). The effects of sequential pictorial cues, self-recording , and praise on the job task sequencing of retarded adults. *Journal of Applied Behavioral Analysis, 12*, 355-361.

Cooper, H. (2010). *Research synthesis and meta-analysis: A step-by-step approach* (4th ed.). Thousand Oaks, CA: SAGE Publications.

Council for Exceptional Children. (2014). *Council for Exceptional Children standards for evidence-based practices in special education.* Retrieved from http://www.cec.sped.org/Standards/Evidence-Based-Practice-Resources-Original

Devlin, P. (2008). Enhancing the job performance of employees with disabilities using the self-determined career development model. *Education and Training in Developmental Disabilities, 43*, 502-513.

DiPipi-Hoy, C., Jitendra, A.K., & Kern, L. (2009). Effects of time management instruction on adolescents' ability to self-manage time in a vocational setting. *The Journal of Special Education, 43,* 145-159. doi: 10.1177/0022466908317791

Dolyniuk, C.A., Kamens, M.W., Corman, H., DiNardo, P.O., Totaro, R.M., & Rockoff, J.C. (2002). Students with developmental disabilities go to college: description of a collaborative transition project on a regular college campus. *Focus Autism Other Developmental Disabilities, 17,* 237-241.

Dotto-Fojut, K.M., Reeve, K.F., Townsend, D.B., & Progar, P.R. (2011). Teaching adolescents with autism to describe a problem and request assistance during simulated vocational tasks. *Research in Autism Spectrum Disorders, 5*, 826-833. doi: 10.1016/j.rasd.2010.09.12

Dunn, O.J. (1964). Multiple contrasts using rank sums. *Technometrics, 6,* 241-252. doi: 10.2307/1266041

Gast, D.L., & Ledford, J.R. (Eds.). (2014). *Single case research methodology: Applications in special education and behavior sciences (2nd ed.)*. New York, NY: Routledge.

Glennon T. (2001). The stress of the university experience with students with asperger syndrome. *Work, 17*, 183–190.

Goh, A.E., & Bambara, L.M. (2013). Video self-modeling: a job skills intervention with individuals with intellectual disability in employment settings. *Education and Training in Autism and Developmental Disabilities, 48*, 103-119.

Grigal, M., & Deschamps, A. (2012) Transition education for adolescents with intellectual disability. In M.L. Wehmeyer & K.W. Webb (Eds.), *Handbook of Adolescent Transition Education for Youth with Disabilities* (pp. 398-416). New York, NY: Routledge

Hanley-Maxwell, C. & Izzo, M.V. (2012). Preparing students for the 21st Century workforce. In M.L. Wehmeyer & K.W. Webb (Eds.), *Handbook of Adolescent Transition Education for Youth with Disabilities* (pp. 139-155). New York, NY: Routledge

Hendricks, D.R., & Wehman, P. (2009). Transition from school to adulthood for youth with autism spectrum disorders. *Focus on Autism and Other Developmental Disabilities, 24*(2), 77-88. doi: 10.1177/1088357608329827

Hendricks, D.R. (2010). Employment and adults with autism spectrum disorders: challenges and strategies for success. *Journal of Vocational Rehabilitation, 32*, 125-134. doi: 10.3233/JVR-2010-0502

Hine, J., & Wolery, M. (2006). Using point-of-view modeling to teach play to preschoolers with autism. *Topics in Early Childhood and Special Education*, *26*, 83–93. doi:10.1177/02711214060260020301

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

Hughes, C. (2001). Transition to adulthood: Supporting young adults to access social, employment, and civic pursuits. *Mental Retardation and Developmental Disabilities Research Reviews, 7*, 84-90.

Hume, K., & Odom, S. (2007). Effects or an individual work system on the independent functioning of students with autism. *Journal of Autism and Developmental Disorders, 37*, 1166-1180. doi: 10.1007/s10803-006-0260-5

Individuals with Disabilities Education Improvement Act (IDEA) of 2004, PL 108-446, 20 U.S.C. ss 1,400 et seq.

Johnson, D.R. (2012). Policy and adolescent transition education. In M.L. Wehmeyer & K.W. Webb (Eds.), *Handbook of Adolescent Transition Education for Youth with Disabilities* (pp. 11-31). New York, NY: Routledge

Kagohara, D.M., Van der Meer, L., Ramdoss, S., O'Reilly, M.F., Lancioni, G.E., Davis, T.N., … Sigafoos, J. (2013). Using iPods ® and iPads ® in teaching programs for individuals with developmental disabilities: A systematic review. *Research in Developmental Disabilities, 34*, 147-156.

Kazdin, A. E. (2011). *Single-case research designs: Methods for clinical and applied settings* (2nd ed.). New York, NY: Oxford University Press.

Kelly, J.A., Wildman, B.G., & Berler, E.S. (1980). Small group behavioral training to improve the job interview skills repertoire of mildly retarded adolescents. *Journal of Applied Behavior Analysis, 13*, 461-471.

Kemp, D.C., & Carr, E.G. (1995). Reduction of severe problem behavior in community employment using an hypothesis-driven multicomponent intervention approach. *Journal of the Association of Persons with Severe Handicaps, 20*(4), 229-247.

Kratochwill, T.R., Hitchcock, J., Horner, R.H., Levin, J.R., Odom, S.L., Rindskopf, D.M., & Shadish, W.R. (2010). Single-case design technical documentation. Retrieved from What Works Clearinghouse website: http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf

Kratochwill, T.R., Hitchcock, J.H., Horner, R.H., Levin, J.R., Odom, S.L., Rindskopf, D.M., Shadish, W.R. (2013). Single-case intervention research design standards. *Remedial and Special Education, 34*(1), 26-38. doi: 10.1177/0741932512452794

Kruskal, W.H., & Wallis, A.W. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association, 47*, 583-621. doi: 10.2307/2280779

Lattimore, L.P., Parsons, M.B., & Reid, D.H. (2006). Enhancing job-site training of supported workers with autism: A reemphasis on simulation. *Journal of Applied Behavior Analysis, 39*, 91-102.

Lattimore, L.P., Parsons, & M.B., Reid, D. (2008). Simulation training of community job skills for adults with autism: A further analysis. *Behavior Analysis in Practice, 1*, 24-29.

Lattimore, L.P., Parsons, & M.B., Reid, D. (2009). Rapid training of community job skill to nonvocal adults with autism: An extension of intensive training. *Behavior Analysis in Practice, 2*, 34-42.

Lee, G.K., & Carter, E.W. (2012). Preparing transition-age students with high-functioning autism spectrum disorders for meaningful work. *Psychology in the Schools, 49*, 988-1000. doi: 10.1002/pits_21651

Likins, M., Salzberg, C.L., Stowitschek, J.J., Lignugaris-Kraft, & Curl, R. (1989). Co-worker implemented job training: the use of coincidental training and quality-control checking on the food preparation skills of trainees with mental retardation. *Journal of Applied Behavioral Analysis, 22*, 381-393.

Lipsey, M.W., & Wilson, D.B. (2001). *Practical meta-analysis.* Applied Social Research Methods Series, Volume 49. Bickman, L., & Rog, D.J. (Eds.). Thousand Oaks, CA: Sage Publications.

Maggin, D.M., Briesch, A.M., & Chafouleas, S.M. (2013). An application of the What Works Clearinghouse Standards for evaluating single-subject synthesis of the self-management literature base. *Remedial and Special Education, 34,* 44-58. doi: 10.1177/0741932511435176

Martin, J.E., Elias-Burger, S., & Mithuag, D.E. (1987). Acquisition and maintenance of time-based task change sequence. *Education and Training in Mental Retardation, 22*(4), 250-255.

Mason, R.A., Davis, H.S., Boles, M.B., & Goodwyn, F. (2013a). Efficacy of point-of-view video modeling: A meta-analysis. *Remedial and Special Education, 34*, 333-345. doi: 10.1177/0741932513486298

Mason, R.A., Ganz, J.B., Parker, R.I., Boles, M.B., Davis, H.S., & Rispoli, M.J. (2013b). Video-based modeling: Differential effects due to treatment protocol. *Research in Autism Spectrum Disorders*, *7*(1), 120-131.

Mason, R.A., Ganz, J.B., Parker, R.I., Burke, M.D., & Camargo, S.P. (2012a). Moderating factors of video-modeling with other as model: A meta-analysis of single-case studies. *Research in Developmental Disabilities, 33*, 1076-1086. doi: 10.1016/j.ridd.2012.01.016

Mason, R.A., Rispoli, M., Ganz, J.B., Boles, M.B., Orr, K. (2012b). Effects of video modeling on communicative social skills of college students with Asperger syndrome. *Developmental Neurorehabilitation, 15*, 425-434. doi: 10.3109/17518423.2012.704530

Matson, J. L., & Rivet, T. T. (2008). Characteristics of challenging behaviors in adults with autistic disorder, PDD-NOS, and intellectual disability. *Journal of Intellectual and Developmental Disability*, *33*(4), 323-329. doi:10.1080/13668250802492600

Mazurek, M. O., & Kanne, S. M. (2010). Friendship and internalizing symptoms among children and adolescents with ASD. *Journal of Autism and Developmental Disorders, 40*(12), 1512-1520. doi: 10.1007/s10803-010-1014-y

McCoy, K., & Hermansen, E. (2007). Video modeling for individuals with autism: A review of model types and effects. *Education and Treatment of Children, 30*, 183-213.

Mechling, L.C., & Ortega-Hurndon, F. (2007). Computer-based video instruction to teach young adults with moderate intellectual disabilities to perform multiple step, job tasks in a generalized setting. *Education and Training in Developmental Disabilities, 42,* 24-37.

Mechling, L.C., & Savidge, E.J. (2011). Using a personal digital assistant to increase completion of novel tasks and independent transitioning by students with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 41*, 687-704. doi: 10.1007/s10803-010-1088-6

Mechling, L.C., & Ayres, K.M. (2012). A comparative study: completion of fine motor office related tasks by high school students with autism using video models on large and small screen sizes. *Journal of Autism and Developmental Disorders.* doi: 10.1007/s10803-012-1484-1

Migliore, A., & Butterworth, J. (2008). Trends in outcomes of the vocational rehabilitation program for adults with developmental disabilities: 1995-2005. *Rehabilitation Counseling Bulletin, 52*, 35-44.

Mitchell, R.J., Schuster, J.W., Collins, B.C., & Gassaway, L.J. (2000). Teaching vocational skills with a faded auditory prompting system. *Education and Training in Mental Retardation and Developmental Disabilities, 35*, 415-427.

Morgan, R.L., & Salzberg, C.L. (1992). Effects of video-assisted training on employment-related social skills of adults with severe mental retardation. *Journal of Applied Behavior Analysis, 25*, 365-383.

National Secondary Transition Technical Assistance Center. (2015). *Evidence-based practices.* Retrieved from http://www.nsttac.org/content/evidence-based-practices

Newman, L., Wagner, M., Cameto, R., & Knokey, A.-M.(2009). *The Post-High School Outcomes of Youth with Disabilities up to 4 Years After High School. A Report of Findings from the National Longitudinal Transition Study-2 (NLTS2) (NCSER 2009-3017).* Menlo Park, CA: SRI International. Available at www.nlts2.org/reports/2009_04/nlts2_report_2009_04_complete.pdf.

Newman, L., Wagner, M., Knokey, A.-M., Marder, C., Nagle, K., Shaver, D., … Schwarting, M. (2011). *The post-high school outcomes of young adults with disabilities up to 8 years after high school: A report from the National Longitudinal Transition Study-2 (NLTS2)* (NCSER 2011-3005). Menlo Park, CA: SRI International. Available at www.nlts2.org/reports/

Ninci, J., Neely, L.C., Hong, E.R., Boles, M.B., Gilliland, W.D., Ganz, J.B., … Vannest, K.J. (2015). Meta-analysis of single-case research on teaching functional living skills to individuals with ASD. *Review Journal of Autism and Developmental Disorders, 2*, 184-198. doi: 10.1007/s40489-014-0046-1

No Child Left Behind (NCLB) Act of 2001, Pub. L. No. 107-110, sec. 115, Stat. 1425 (2002).

Norman, J. M., Collins, B., & Schuster, J. (2001). Using an instructional package including video technology to teach self-help skills to elementary students with mental disabilities. *Journal of Special Education Technology*, *16*, 5–18.

Palmen, A., Didden, R., & Lang, R. (2012). A systematic review of behavioral intervention research in adaptive skill building in high functioning young adults with autism spectrum disorder. *Research in Autism Spectrum Disorders, 6,* 602-617.

Palmer, R. (2014). Technical and vocational skills and post-2015: Avoiding another vague skills goal? *International Journal of Educational Development, 39*, 32-39.

Parker, R.I., & Vannest, K.J. (2009). An improved effect size for single-case research: Nonoverlap of all pairs. *Behavior Therapy, 40*, 357-367.

Parker, R.I., Vannest, K.J., & Davis, J.L. (2011a). Effect size in single-case research: A review of nine nonoverlap techniques. *Behavior Modification, 35*, 303-322. doi: 10.1177/0145445511399147

Parker, R.I., Vannest, K.J., Davis, J.L., & Sauber, S.B. (2011b). Combining nonoverlap and trend for single-case research: Tau-U. *Behavior Therapy, 42*, 284-299. doi: 10.1016/j.beth.2010.08.006

Parson, M.B., Reid, D.H., Green, C.W., Browning, L.B. (1999). Reducing individualized job coach assistance provided to persons with multiple severe disabilities in supported work. *Research and Practice for Persons with Severe Disabilities, 24*, 292-297. doi: 10.2511/rpsd.24.4.292

Rai, K. (2008). Technology to teach self-help skills to elementary students with mental disabilities. *Journal of the Indian Academy of Applied Psychology, 34*, 201-214.

Ratto, A. B., Turner-Brown, L., Rupp, B. M., Mesibov, G. B., & Penn, D. L. (2010). Development of the contextual assessment of social skills (CASS): A role play measure of social skill for individuals with high-functioning autism. *Journal of Autism and Developmental Disorders, 41*(9), 1277-1286. doi: 10.1007/s10803-010-1147-z

Rayner, C., Denholm, C., & Sigafoos, J. (2009). Video-based intervention for individuals with Autism: Key questions that remain unanswered. *Research in Autism Spectrum Disorders*, *3*, 291–303. doi:10.1016/j.rasd.2008.09.001

Reichow, B., Volkmar, F.R., Cicchetti, D.V. (2008). Development of the evaluative method for evaluating evidence-based practices in autism. *Journal of Autism and Developmental Disorders, 38*, 1311-1319. doi: 10.1007/s10803-007-0517-7

Riffel, L.A., Wehmeyer, M.L., Turnbull, A.P., Lattimore, J., Davies, D., Stock, S., & Fisher, S. (2005). Promoting independent performance of transition-related tasks

using a palmtop PC-based self-directed visual and auditory prompting system. *Journal of Special Education Technology, 20*(2), 5-14.

Roth, M.E., Gillis, J.M., & Reed, F.D. (2014). A meta-analysis of behavioral interventions for adolescents and adults with autism spectrum disorders. *Journal of Behavioral Education, 23*, 258-286. doi: 10.1007/s10864-013-9189-x

Rusch, F.R., & Dattilo, J. (2012). Employment and self-management: A meta-evaluation of seven literature reviews. *Intellectual and Developmental Disabilities, 50*(1), 69-75. doi: 10.1352/1934-9556-50.1.69

Sanford, C., Newman, L., Wagner, M., Cameto, R., Knokey, A.-M., & Shaver, D. (2011). *The Post-High School Outcomes of Young Adults With Disabilities up to 6 Years After High School. Key Findings From the National Longitudinal Transition Study-2 (NLTS2)* (NCSER 2011-3004). Menlo Park, CA: SRI International. Available at www.nlts2.org/reports/

Schreibman, L., Whalen, C., & Stahmer, A. (2000). The use of video priming to reduce disruptive transition behavior in children with autism. *Journal of Positive Behavior Interventions*, *2*, 3–11.

Shipley-Benamou, R., Lutzker, J. R., & Taubman, M. (2002). Teaching daily living skills to children with autism through instructional video modeling. *Journal of Positive Behavior Interventions*, *4*, 165–175. doi:10.1177/10983007020040030501

Shukla-Mehta, S., Miller, T., & Callahan, K. J. (2010). Evaluating the effectiveness of video instruction on social and communication skills training for children with

autism spectrum disorders: A review of the literature. *Focus on Autism and Other Developmental Disabilities*, *25*, 23–36. doi:10.1177/1088357609352901

Sigafoos, J., O'Reilly, M., Cannella, H., Upadhyaya, M., Edrisinha, C., Lancioni, G. E., & Young, D. (2005). Computer-presented video prompting for teaching micro-wave oven use to three adults with developmental disabilities. *Journal of Behavioral Education*, *14*, 189–201. doi:10.1007/s10864-005-6297-2

Soares, Denise (2011). Effect Size and Moderators of Effects for Token Economy Interventions. Doctoral dissertation, Texas A&M University. Available electronically from http ://hdl.handle.net /1969.1/ETD-TAMU-2011-12-10436

Taylor, J.L., McPheeters, M.L., Sathe, N.A., Dove, D., Veenstra-VanderWeele, J., & Warren, Z. (2012). *Pediatrics*, 130, 531-538. doi: 10.1542/peds.2012-0682

Test, D.W., Richter, S., Walker, A.R. (2012) Life skills and community-based instruction. In M.L. Wehmeyer & K.W. Webb (Eds.), *Handbook of Adolescent Transition Education for Youth with Disabilities* (pp. 121-138). New York, NY: Routledge

Tetreault, A. S., & Lerman, D. C. (2010). Teaching social skills to children with autism using point-of-view video modeling. *Education & Treatment of Children*, *33*, 395–419.

Van Laarhoven, T., Van Laarhoven-Myers, T., & Zurita, L.M. (2007). The effectiveness of using a pocket PC as a video modeling and feedback device for individuals with developmental disabilities in vocational settings. *Assistive Technology Outcomes and Benefits, 4*(1), 28-45.

Van Laarhoven, T., Zurita, L. M., Johnson, J. W., Grider, K. M., & Grider, K. L. (2009). Comparison of self, other, and subjective video models for teaching daily living skills to individuals with developmental disabilities. *Education and Training in Developmental Disabilities*, *44*, 509–522.

Vannest, K.J., Parker, R.I., & Gonen, O. (2011). Single Case Research: web based calculators for SCR analysis. (Version 1.0) [Web-based application]. Available from singlecaseresearch.org

Wacker, D.P., Berg, W.K., Berrie, P., & Swatta, P. (1985). Generalization and maintenance of complex skills by severely handicapped adolescents following picture prompt training. *Journal of Applied Behavior Analysis, 18,* 329-226.

Wacker, D.P., Berg, W.K., Choisser, L., & Smith, J. (1989). Evaluation of the generalize effects of a peer-training procedure with moderately retarded adolescents. *Journal of Applied Behavior Analysis, 22*, 261-273.

Walker, A.R., Richter, S., Uphold, N.M., Test, D.W. (2010). Review of the literature on community-based instruction across grade-levels. *Education and Training in Autism and Developmental Disabilities, 45*(2), 242-267.

Wehmeyer, M.L. (1994). Employment status and perceptions of control of adults with cognitive and developmental disabilities. *Research in Developmental Disabilities*, *15*(2), 119-131.

Wehmeyer, M.L., Palmer, S.B., Smith, S.J., Parent, W., Davies, D.K., & Stock, S. (2006). Technology use by people with intellectual and developmental

disabilities to support employment activities: A single-subject design meta analysis. *Journal of Vocational Rehabilitation, 24*, 81-86.

Wehmeyer, M.L., & Webb, K.W. (2012). An introduction to adolescent transition education. In Wehmeyer, M.L. & Webb, K.W. (Eds.). *Handbook of adolescent transition education for youth with disabilities* (pp. 3-10). New York, NY: Routledge.

What Work Requires of Schools: A SCANS Report for America 2000 (1991). The Secretary's Commission on Achieving Necessary Skills. US Department of Labor.

Wilcoxon, F., Rhodes, L.J., & Bradley, R.A. (1963). Two sequential two-sample grouped rank tests with applications to screening experiments. *Biometrics, 19*, 58-77.

Wolery, M. (2013). A commentary: Single-case design technical document of the What Works Clearinghouse. *Remedial and Special Education, 43*, 39-43. doi: 10.1177/0741932512468038

Wolf, M.M. (1978). Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis, 11,* 203-214.

Wood, J. J., Drahota, A., Sze, K., Har, K., Chiu, A., & Langer, D. A. (2009). Cognitive behavioral therapy for anxiety in children with autism spectrum disorders: A randomized, controlled trial. *Journal of Child Psychology and Psychiatry, 50*, 224-234. doi: 10.1111/j.1469-7610.2008.01948.x

APPENDIX A

EMPLOYMENT LITERATURE SEARCH MATRIX AND INCLUSION SCREENING RESULTS

Table A1

Developmental Disability Employment Literature Search Matrix

"AND"	employ*	career*	vocation*	employ*skill*	career*skill*	vocation* skill*	job skill*
autis*	940	156	236	21	0	30	35
Asperger*	119	25	25	4	0	5	2
ASD*	428	53	68	10	0	9	18
PDD*	114	1	0	0	0	1	0
Pervasive developmental disorder	135	10	21	6	0	7	11
development* disab*	1770	322	503	29	6	37	71
low-incidence dis*	4	8	4	0	0	0	0
intellectual* disab*	1239	197	328	16	0	15	36
mental* retard*	1915	375	813	39	10	49	138
multipl disab*	705	148	246	4	4	4	13

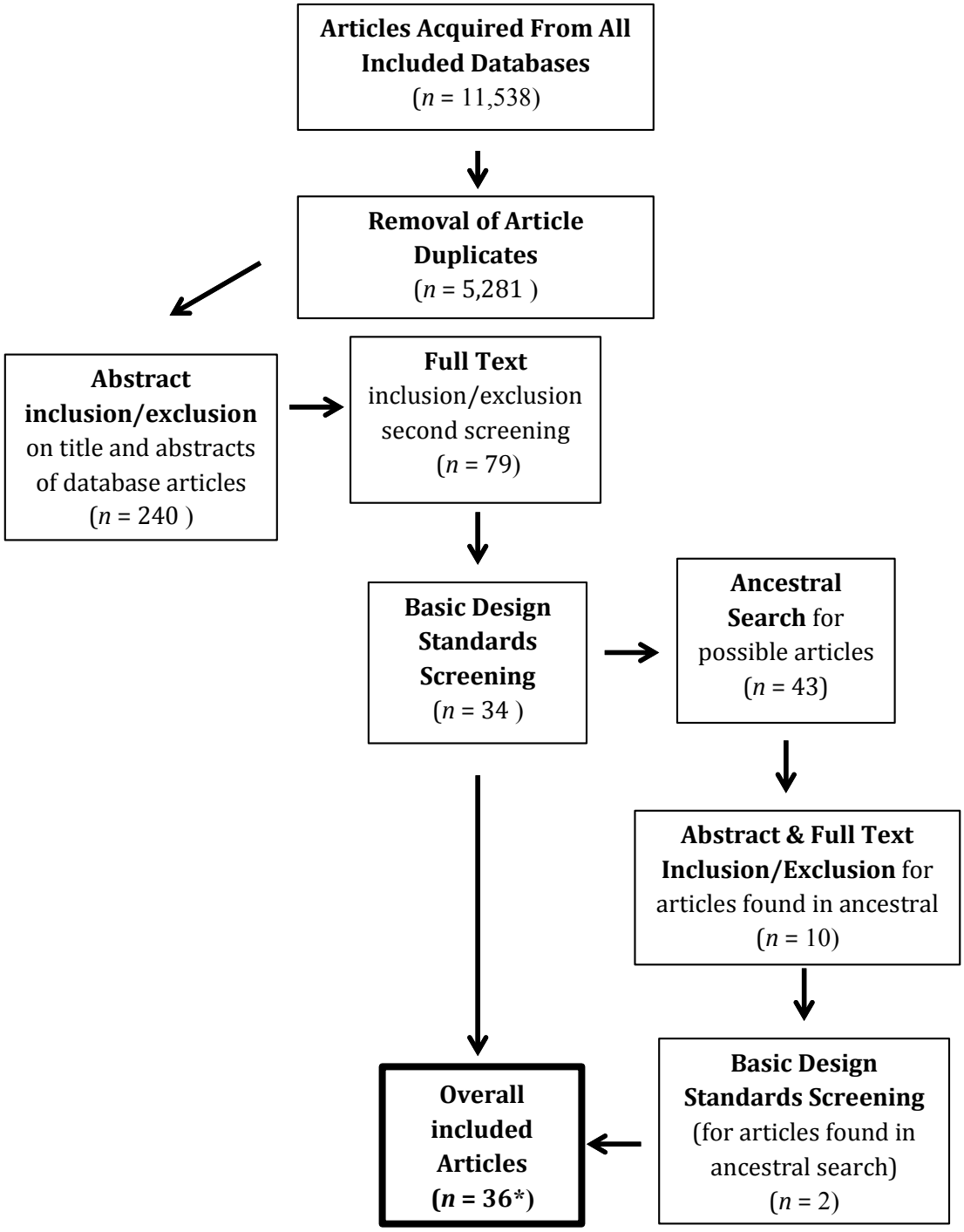

Figure A1. A summary of the total number of sources and articles remaining after each inclusion and exclusion screening

Note. *Three articles included two studies in each article for a total of 39 studies. Only 38 of these studies were unique (i.e., at least one different participant)

APPENDIX B

MODERATOR CODES AND DEFINITIONS OF DESCRIPTIVE STUDY CHARACTERISTICS

Participant	Gender	Age	IQ	Diagnosis	Setting	Intervention	Interventionist	Task Complexity	Target Behavior
Please put the name of the participant	**M** = Male	**ELEM** = Ages less than 12	**HIGH** = IQ scores of 50 or more (Mild intellectual disability or no intellectual disability)	**ASD** = Autism, Asperger's, PDD-NOS, Autism Spectrum Disorders	**SIM** = Simulation, any setting where skills are being taught to use in a place that is not a place of employment or community setting where the skills would naturally occur (i.e. the classroom, the school office, etc.)	**VM**=Video modeling includes an individual watching a video or live demonstration (in-vivo modeling) of an individual (adult, peer, self, point of view, etc.) performing/modeling the targeted skill	**TCHR** = teacher or instructor	**LOW** = low complexity (1-10 steps in task analysis for each task)	**CLEAN** = cleaning a work space, or cleaning a certain object or area for clients ***Job Specific Skill**
	F = Female	**MS/HS** = Ages 12-15	**LOW** = IQ scores of less than 50 (Moderate to severe intellectual disability)	**ASDID** = ASD with Intellectual Disability	**NAT** = Natural setting where the skills being taught would naturally occur (i.e. place of employment or community setting)	**AC/AP** = Audio cueing/audio coaching includes pre-recorded or live prompts via a teacher/trainer/interventionist or device (ear piece) usually from a distance while the participant performs the task	**JC** = Job Coach	**MOD** = moderate complexity (11-20 steps in task analysis for each task)	**ASSEM** = Assembly of packaging, stuffing folders/envelopes, putting together more than one piece of an object ***Job Specific Skill**
		HS = Ages 16-21		**ID** = Intellectual Disability		**VIS** = Visuals include static pictures or words in a single form or multiple pictures or words modeling the targeted behavior for each step of a task analysis.	**RSCHR** = researcher or experimenter	**HIGH** = high complexity (21 or more steps in task analysis for each task)	**REP** = Repetitive tasks that do not require assembling (see **ASSEM** code) and happen more than once in a row (i.e. copy, fax, stocking, doing a task with one object each time) ***Job Specific Skill**
		ADULT = Ages 22 and over				**PRMT** = Most- to-least OR Least-to-most prompting includes the systematic increase or decrease of applying more or less intrusice prompts systematic use of more intrusive prompts.	**OTH** = Other interventionists not specified or match with codes above	**SOC** = social skill that involves communicating with others regarding the task (i.e. ask for help, interview with an employer, interacting with customers)	**SS** = social skills like requests, interviewing, interacting with customers or asking questions ***Employability Skills**
								TRANS = Transition from tasks or tasks steps	**TRANS** = Transition from tasks or tasks steps ***Employability Skills**
						OTH = Other interventions that are not specified, do not match with codes above, or contain more than one of the codes above		**OTH** = Other task steps not specified or match with codes above	

APPENDIX C

TAU-U AND SIGNIFICANCE OUTCOMES

Table C1
Tau-U Scores and Significant Differences for Intervention Type Moderator

Intervention Tau-U Scores				
Intervention Type	Number of Studies	Number of Participants	Number of Contrasts	Tau-U [CI95]
VM (video modeling)	15	44	112	0.83 [0.79, 0.87]
AC (audio cueing/coaching)	4	13	17	0.97 [0.85, 1.00]
VIS (visuals)	8	31	51	0.97 [0.91, 1.00]
PRMT (prompting)	5	14	28	0.93 [0.84, 1.00]

Intervention Dunn Post-Hoc Results			
Group Comparisons	Difference in average ranks	Cutoff at alpha = 0.05	Significant Differences = **
AC-PRMT	18.83	48.82	
AC-VIS	30.11	44.47	
AC-VM	52.37	41.33	**
PRMT-VIS	11.28	37.35	
PRMT-VM	33.54	33.55	
VIS-VM	22.26	26.82	

Table C2

Video Modeling: Tau-U Scores and Significant Differences for Target Behaviors

Target Behavior Tau-U Scores				
Target Behaviors	**Number of Studies**	**Number of Participants**	**Number of Contrasts**	**Tau-U [CI95]**
REP (repetitive)	5	15	36	0.69 [0.61, 0.78]
ASSEM (assembly)	5	12	21	0.89 [0.80, 0.98]
CLEAN (cleaning)	2	3	5	0.92 [0.76, 1.00]
SS (social skills)	6	17	31	0.85 [0.78, 0.92]
TRANS (transition between tasks/steps)	2	5	16	0.88 [0.75, 1.00]
Target Behavior Dunn Post-Hoc Results				
Group Comparisons	**Difference in average ranks**		**Cutoff at alpha = 0.05**	**Significant Differences = ****
ASSM-CLEAN	3.67		44.15	
ASSEM-REP	27.04		24.36	**
ASSEM-SS	9.17		25.08	
ASSEM-TRANS	0.40		29.44	
CLEAN-REP	30.71		42.35	
CLEAN-SS	12.84		42.76	
CLEAN-TRANS	4.07		45.46	
REP-SS	17.87		21.74	
REP-TRANS	26.64		26.66	
SS-TRANS	8.7732		27.3133	

Table C3

Video Modeling: Tau-U Scores and Significant Differences for Task Complexity

Task Complexity Tau-U Scores				
Task Complexity	**Number of Studies**	**Number of Participants**	**Number of Contrasts**	**Tau-U [CI95]**
LOW (1-10steps)	4	11	53	0.84 [0.73, 0.95]
MOD (11-20 steps)	4	12	19	0.71 [0.62, 0.80]
HIGH (21 steps and over)	1	4	4	0.77 [0.52, 1.00]
SOC (complex social skills)	6	17	31	0.85 [0.78, 0.92]
TRANS (lower complexity transition skills)	2	5	16	0.88 [0.75, 1.00]
Task Complexity Dunn Post-Hoc Results				
Group Comparisons	**Difference in average ranks**		**Cutoff at alpha = 0.05**	**Significant Differences = ****
HIGH-LOW	17.95		41.67	
HIGH-MOD	5.93		42.57	
HIGH-SOC	19.00		41.11	
HIGH-TRANS	26.06		43.26	
LOW-MOD	23.87		23.55	**
LOW-SOC	1.05		20.80	
LOW-TRANS	8.12		24.78	
MOD-SOC	24.92		22.55	**
MOD-TRAN	31.99		26.26	**
SOC-TRAN	7.07		23.82	

Table C4

Video Modeling: Tau-U Scores for Participant Characteristic Moderators

	Number of Studies	Number of Participants	Number of Contrasts	Tau-U [CI_{95}]
Age				
MS/HS (12-15yrs.)	3	7	31	0.85 [0.75, 0.96]
HS (16-21yrs.)	9	23	57	0.82 [0.77, 0.88]
ADULT (22yrs. & over)	7	14	24	0.83 [0.75, 0.90]
IQ				
HIGH (50 or above)	9	16	43	0.87 [0.80, 0.94]
LOW (below 50)	9	15	44	0.92 [0.85, 0.99]
Diagnosis				
ASD (autism spectrum disorder)	3	10	10	0.79 [0.68, 0.89]
ASDID (ASD + ID)	5	12	45	0.81 [0.73, 0.90]
ID (intellectual disability)	9	22	57	0.85 [0.80, 0.91]

Table C5

Video Modeling: Tau-U Scores for Setting and Interventionist Moderators

	Number of Studies	Number of Participants	Number of Contrasts	Tau-U [CI95]
Setting				
NAT (natural)	8	24	44	0.81 [0.75, 0.87]
SIM (simulation)	8	24	68	0.85 [0.79, 0.91]
Interventionist				
JC (job coach)	1	4	12	0.65 [0.52, 0.79]
RSCHR (researcher)	2	6	14	0.92 [0.79, 1.00]
TCHR (teacher)	1	3	21	0.78 [0.64, 0.92]

APPENDIX D

QUALITY INDICATOR OUTCOMES

Table D1

Basic Design Standards

	Study Name (Date)	IV/Intervention	IOA			Effect	Data Points	OVERALL SCORE
		Purposeful Manipulation of IV	IOA Recorded	IOA Session Totals	IOA Percentage Agreement	Attempts To Present Effect	Data Points in Each Phase	Meeting Design Standards
Studies that Meet Basic Design Standards	Cihak, Alberto, Kessler, and Taber (2004)	2	2	2	2	2	2	2
	Cihak, Kessler, and Alberto (2007)	2	2	2	2	2	2	2
	Cihak, Kessler, and Alberto (2008)	2	2	2	2	2	2	2
	Dotto-Fojut, Reeve, Townsend, and Progar (2011)	2	2	2	2	2	2	2
	Kemp and Carr (1995)	2	2	2	2	2	2	2
Studies that Meet Basic Design Standards with Reservations	Allen, Wallace, Greene, Bowen, and Burke (2010a)	2	2	1	1	2	2	1
	Allen, Wallace, Renes, Bowen, and Burke (2010b)	2	2	1	2	2	2	1
	Allen, Burke, Howard, Wallace, and Bowen (2012)	2	2	1	1	2	1	1
	Bennett, Brady, Scott, Dukes, and Frain (2010)	2	2	1	1	2	1	1
	Bennett, Rangasamy, and Honsberger (2013a)	2	2	1	1	2	2	1
	Bennett, Rangasamy, and Honsberger (2013b)	2	2	1	1	2	2	1
	Bereznak, Ayres, Mechling, and Alexander (2012)	2	2	1	1	2	1	1
	Cavkaytar (2012)	2	2	1	1	2	1	1

	Study Name (Date)	IV/Intervention	IOA			Effect	Data Points	OVERALL SCORE
		Purposeful Manipulation of IV	IOA Recorded	IOA Session Totals	IOA Percentage Agreement	Attempts To Present Effect	Data Points in Each Phase	Meeting Design Standards
Studies that Meet Basic Design Standards with Reservation	Chandler, Schuster, and Stevens (1993)	2	2	2	2	2	1	1
	Chang, Kang, & Huang (2013)	2	2	1	1	2	1	1
	Connis (1997)	2	2	1	2	2	2	1
	Devlin (2008)	2	2	1	2	2	2	1
	DiPipi-Hoy, Jitendra, and Kern. (2009)	2	2	1	1	2	2	1
	Goh and Bambara (2013)	2	2	1	2	2	1	1
	Hume and Odom (2007)	2	2	1	2	2	1	1
	Kelly, Wildman, and Berler (1980)	2	2	1	2	2	1	1
	Lattimore, Parsons, and Reid (2006)	2	2	2	1	2	1	1
	Lattimore, Parsons, and Reid (2008) Study 1*	2	2	2	1	1	1	1
	Lattimore, Parsons, and Reid (2008) Study 2*	2	2	2	1	2	1	1
	Lattimore, Parsons, and Reid (2009)	2	2	2	1	2	1	1
	Likins, Salzberg, Stowitschek, Lignugaris/Kraft, & Curl (1989) Study 1*	2	2	1	1	2	2	1
	Likins et al. (1989) Study 2*	2	2	1	1	2	2	1

	Study Name (Date)	IV/Intervention	IOA			Effect	Data Points	OVERALL SCORE
		Purposeful Manipulation of IV	IOA Recorded	IOA Session Totals	IOA Percentage Agreement	Attempts To Present Effect	Data Points in Each Phase	Meeting Design Standards
Studies that Meet Basic Design Standards with Reservation	Martin et al. (1987)	2	2	1	1	2	1	1
	Mechling and Ortega-Hurndon (2007)	2	2	1	1	2	1	1
	Mechling and Savidge (2011)	2	2	1	1	2	1	1
	Mechling and Ayres (2012)	2	2	2	1	2	1	1
	Mitchell, Schuster, Collins, and Gassaway (2000)	2	2	1	1	2	1	1
	Morgan and Salzberg (1992) Study 1*	2	2	1	2	2	1	1
	Morgan and Salzberg (1992) Study 2*	2	2	1	2	2	1	1
	Parson, Reid, Green, and Browning (1999)	2	2	2	1	2	2	1
	Riffel et al. (2005)	2	2	1	1	2	2	1
	Van Laarhoven, Van Laarhoven-Meyers, and Zurita (2007)	2	2	2	2	2	1	1
	Wacker, Berg, Berrie, & Swatta (1985)	2	2	2	1	2	2	1
	Wacker, Berg, Choisser, and Smith (1989)	2	2	1	1	2	1	1

Note. *2 studies were included in one article

Table D2

Descriptive Quality Standards

	Study Name (Date)	Participant Description	Setting Description	Interventionist Description	Baseline and Intervention Procedure Description	Dependent Variable Description	Overall Score
Sufficient Description	Dotto-Fojut et al. (2011)	2	2	2	2	2	2
	Likins et al. (1989) Study 1*	2	2	2	2	2	2
Minimal Description	Bennett et al. (2010)	2	1	2	1	2	1
	Bennett et al. (2013a)	2	1	1	1	2	1
	Bennett et al. (2013b)	2	1	1	2	2	1
	Bereznak et al. (2012)	2	2	1	2	2	1
	Cavkaytar (2012)	2	1	1	1	2	1
	Chandler et al. (1993)	1	2	1	2	1	1
	Cihak et al. (2004)	2	1	1	2	2	1
	Cihak et al. (2007)	1	1	1	1	2	1
	Cihak et al. (2008)	1	1	1	1	1	1
	Devlin (2008)	1	1	1	1	1	1
	DiPipi-Hoy et al. (2009)	2	1	1	2	2	1
	Goh and Bambara (2013)	2	1	1	2	2	1
	Kemp and Carr (1995)	2	1	2	2	2	1
	Lattimore et al. (2006)	2	1	2	2	1	1
	Lattimore et al. (2009)	2	1	2	1	1	1

	Study Name (Date)	Participant Description	Setting Description	Interventionist Description	Baseline and Intervention Procedure Description	Dependent Variable Description	Overall Score
Minimal Description	Likins et al. (1989) Study 2*	2	2	2	1	1	1
	Mechling and Ortega-Hurndon (2007)	2	2	1	2	2	1
	Mechling and Savidge (2011)	2	2	1	2	2	1
	Mechling and Ayres (2012)	2	1	1	2	2	1
	Mitchell et al. (2000)	2	1	1	2	2	1
	Parson et al. (1999)	2	2	2	1	1	1
Insufficient Description	Allen et al. (2010a)	2	2	0	2	2	0
	Allen et al. (2010b)	2	0	0	2	2	0
	Allen et al. (2012)	2	2	0	2	2	0
	Chang et al. (2013)	1	1	1	1	0	0
	Connis (1997)	0	0	1	2	1	0
	Hume and Odom (2007)	2	1	0	1	1	0
	Kelly et al. (1980)	2	0	0	0	1	0
	Lattimore et al. (2008) Study 1*	2	0	2	2	2	0
	Lattimore et al. (2008) Study 2*	2	0	2	2	2	0
	Martin et al. (1987)	1	0	1	1	1	0
	Morgan and Salzberg (1992) Study 1*	2	0	1	2	2	0
	Morgan and Salzberg (1992) Study 2*	2	0	1	2	2	0
	Riffel et al. (2005)	2	0	1	1	1	0
	Van Laarhoven et al. (2007)	2	2	0	2	2	0
	Wacker et al. (1985)	1	0	0	1	1	0
	Wacker et al. (1989)	1	0	2	1	1	0

Note. *2 studies were included in one article

Table D3
Additional Phase Quality Standards

	Study Name (Date)	Maintenance	Generalization	Fidelity	Social Validity	Overall Score
Minimal Measure	Allen et al. (2010a)	1	1	0	1	1
	Allen et al. (2012)	0	1	1	1	1
	Cavkaytar (2012)	1	1	2	2	1
	Chandler et al. (1993)	2	1	2	0	1
	Cihak et al. (2004)	1	1	2	0	1
	Devlin (2008)	1	0	1	2	1
	DiPipi-Hoy et al. (2009)	0	2	2	2	1
	Dotto-Fojut et al. (2011)	1	1	1	0	1
	Hume and Odom (2007)	1	0	1	1	1
	Kemp and Carr (1995)	1	2	0	1	1
	Lattimore et al. (2009)	2	1	0	1	1
	Mechling and Ortega-Hurndon (2007)	1	2	2	0	1
	Mechling and Savidge (2011)	1	1	2	1	1
	Mitchell et al. (2000)	1	1	2	0	1
	Morgan and Salzberg (1992) Study 1*	1	2	1	0	1
	Van Laarhoven et al. (2007)	1	0	1	1	1
	Wacker et al. (1989)	1	1	1	0	1

	Study Name (Date)	Maintenance	Generalization	Fidelity	Social Validity	Overall Score
Insufficient Measure	Allen et al. (2010b)	1	0	0	1	0
	Bennett et al. (2010)	2	0	1	0	0
	Bennett et al. (2013a)	1	0	2	0	0
	Bennett et al. (2013b)	1	1	2	0	0
	Bereznak et al. (2012)	1	0	2	0	0
	Chang et al. (2013)	1	0	0	1	0
	Cihak et al. (2007)	1	0	2	0	0
	Cihak et al. (2008)	1	0	2	0	0
	Connis (1997)	2	0	0	0	0
	Goh and Bambara (2013)	1	0	1	0	0
	Kelly et al. (1980)	0	1	0	1	0
	Lattimore et al. (2006)	0	1	0	0	0
	Lattimore et al. (2008) Study 1*	1	1	0	0	0
	Lattimore et al. (2008) Study 2*	1	1	0	1	0
	Likins et al. (1989) Study 1*	1	0	0	1	0
	Likins et al. (1989) Study 2*	0	0	0	1	0
	Martin et al. (1987)	0	0	0	0	0
	Mechling and Ayres (2012)	0	0	2	1	0
	Morgan and Salzberg (1992) Study 2*	1	2	0	0	0
	Parson et al. (1999)	1	0	2	0	0
	Riffel et al. (2005)	0	0	0	2	0
	Wacker et al. (1985)	2	1	0	0	0

Note. *2 studies were included in one article

Table D4

Evidence Quality Indicators

Intervention Type	Study Name (Date)	Indicator #1				Indicator #2				Indicator #3							Indicator #4			
		Baseline Analysis				Within Phase Analysis				Between Phase Basic Effects							Overall Effectiveness (All Phases)			
		BA-CH	BA-PR	BA-CV	BA-TR	WP-DP	WP-PR	WP-CV	WP-TR	BW-BE	BW-IL	BW-IT	BW-CL	BW-CV	BW-OV	BW-SP	OV-DP	OV-TE	OV-ER	OV-EE
Video Modeling	Allen et al. (2010a)	1	1	1	1	2	0	0	0	1	1	1	1	1	1		2	2	2	2
	Allen et al. (2010b)	1	1	1	1	2	0	0	0	1	0	0	1	1	0		2	2	2	2
	Morgan and Salzberg (1992) Study 1*	1	1	1	1	2	0	0	1	1	1	1	1	1	1		2	2	2	2
		1	0	0	0	2	1	0	0	1	1	0	1	1	1		2	2	2	2
	Likins et al. (1989) Study 1*	1	1	1	1	2	1	1	1	1	1	1	1	1	1		2	2	2	2
	Likins et al. (1989) Study 2*	1	1	1	0	2	1	1	1	1	0	1	1	1	1		2	2	2	2
	Bereznak et al. (2012)	1	1	1	1	1	0	0	1	1	1	0	1	1	1		1	2	2	1
	Chandler et al. (1993)	1	1	1	0	1	1	1	1	1	0	1	1	1	1		1	2	2	1
		1	1	1	0	1	1	1	1	1	0	1	1	1	1		1	2	2	1
		1	1	1	0	1	1	0	1	1	0	1	1	1	1		1	2	2	1
		1	1	1	0	1	1	1	1	1	0	0	0	1	1		1	2	2	1
	Kelly et al. (1980)	1	1	1	0	1	1	0	0	1	1	0	1	1	1		1	2	2	1
		1	1	1	1	1	0	0	0	1	1	1	1	1	1		1	2	2	1
		1	1	1	1	1	0	0	0	1	0	0	1	1	1		1	2	2	1
	Mechling and Ortega-Hurndon (2007)	1	1	1	0	1	1	1	1	1	0	1	1	1	1		1	2	2	1
		1	1	1	1	1	1	1	1	1	0	1	1	1	1		1	2	2	1
		1	1	1	1	1	1	1	1	1	0	1	1	1	1		1	2	2	1
	Mechling and Saavidge (2011)	1	0	0	0	1	1	1	1	1	1	0	1	1	1		1	2	2	1
		0	0	0	1	1	1	1	1	1	1	0	1	1	1		1	2	2	1
		1	1	1	1	1	1	1	1	1	1	0	1	0	1		1	2	2	1
		1	1	1	1	1	1	1	1	1	1	0	1	0	1		1	2	2	1
		1	1	1	1	1	1	1	1	1	1	1	1	1	1		1	2	2	1

Intervention Type	Study Name (Date)	Indicator #1				Indicator #2				Indicator #3							Indicator #4			
		Baseline Analysis				Within Phase Analysis				Between Phase Basic Effects							Overall Effectiveness (All Phases)			
		BA-CH	BA-PR	BA-CV	BA-TR	WP-DP	WP-PR	WP-CV	WP-TR	BW-BE	BW-IL	BW-IT	BW-CL	BW-CV	BW-OV	BW-SP	OV-DP	OV-TE	OV-ER	OV-EE
Video Modeling	Mechling and Ayres (2012)	1	1	1	1	1	1	1	0	1	1	0	1	1	1		1	2	2	1
		1	1	1	1	1	1	0	1	1	1	1	1	1	1		1	2	2	1
		1	1	1	1	1	1	1	1	1	1	1	1	1	1		1	2	2	1
		1	1	1	1	1	1	1	1	1	1	1	1	1	1		1	2	2	1
		1	1	1	1	1	1	1	1	1	1	1	1	1	1		1	2	2	1
		1	1	1	1	1	1	1	0	1	1	1	1	1	1		1	2	2	1
		1	1	1	1	1	1	1	1	1	1	1	1	1	1		1	2	2	1
		1	1	1	1	1	1	1	1	1	1	1	1	1	1		1	2	2	1
	Morgan and Salzberg (1992) Study 1*	1	1	1	0	1	1	1	0	1	1	0	1	0	0		1	2	2	1
	Morgan and Salzberg (1992) Study 2*	1	0	0	0	1	1	1	1	1	1	1	1	1	1		1	2	2	1
	Van Laarhoven et al. (2007)	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1	2	2	1
		1	1	1	0	1	1	0	1	1	1	1	1	1	1		1	2	2	1
		0	1	1	0	1	1	1	1	1	1	1	1	1	1		1	2	2	1
		1	1	0	0	1	1	0	1	1	1	1	1	1	1		1	2	2	1
AC	Bennett et al. (2013a)	1	1	1	0	2	1	1	1	1	1	0	1	1	1		2	2	2	2
	Bennett et al. (2013b)	1	1	1	1	2	1	1	1	1	1	1	1	1	1		2	2	2	2
		1	1	1	1	2	1	1	1	1	1	1	1	1	1		2	2	2	2
	Allen et al. (2012)	1	1	1	0	1	1	0	1	1	1	1	1	1	1		1	2	2	1
	Bennett et al. (2010)	0	1	1	0	1	1	1	1	0	0	0	1	0	0		1	2	1	1

Intervention Type	Study Name (Date)	Indicator #1				Indicator #2				Indicator #3							Indicator #4			
		Baseline Analysis				Within Phase Analysis				Between Phase Basic Effects							Overall Effectiveness (All Phases)			
		BA-CH	BA-PR	BA-CV	BA-TR	WP-DP	WP-PR	WP-CV	WP-TR	BW-BE	BW-IL	BW-IT	BW-CL	BW-CV	BW-OV	BW-SP	OV-DP	OV-TE	OV-ER	OV-EE
VIS	Cihak et al. (2004)	1	0	1	0	2	1	1	1	1	1	1	1	1	1		2	2	2	2
		1	1	0	1	2	1	1	1	1	1	1	1	1	1		2	2	2	2
	Cihak et al. (2007)	1	1	1	1	2	1	1	1	1	1	1	1	1	1		2	2	2	2
		1	1	1	1	2	1	1	1	1	1	1	1	1	1		2	2	2	2
		1	1	1	1	2	1	1	1	1	1	1	1	1	1		2	2	2	2
		1	1	1	0	2	1	1	1	1	1	1	1	1	1		2	2	2	2
	Cihak et al. (2008)	1	1	1	1	2	1	1	1	1	1	1	1	1	1		2	2	2	2
	Connis (1997)	0	0	1	0	2	1	1	1	1	1	0	1	1	1		2	2	2	2
	Devlin (2008)	1	1	1	0	2	1	1	1	1	1	1	1	1	1		2	2	2	2
	Dotto-Fojut et al. (2011)	1	1	1	1	2	1	0	0	1	1	1	1	0	1		2	2	2	2
		1	1	1	1	2	1	0	1	1	0	0	1	1	1		2	2	2	2
	Martin et al. (1987)	1	0	0	0	2	1	1	1	1	1	1	1	1	1	1	2	2	2	2
		1	0	0	0	1	1	1	1	1	0	0	1	0	0		1	2	1	1
		1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	1
		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	1
	Riffel et al. (2005)	0	0	0	0	2	0	0	0	1	0	0	1	0	0		2	1	1	1
PRMT	Parson et al. (1999)	1	1	1	0	2	1	0	1	1	1	1	1	1	1		2	2	2	2
	Lattimore et al. (2006)	1	1	0	0	1	1	1	1	1	1	1	1	1	1		1	2	2	1
		1	0	1	0	1	1	1	1	1	0	0	1	1	0		1	2	2	1
	Lattimore et al. (2008) Study 1*	1	1	1	0	1	1	1	1	1	1	1	1	1	1		1	2	2	1
	Lattimore et al. (2008) Study 2*	1	1	0	0	1	1	1	0	1	1	1	1	1	1		1	2	2	1
		1	1	1	1	2	1	1	1	1	0	0	1	1	1		2	2	2	2
	Lattimore et al. (2009)	1	1	1	0	1	1	1	1	1	1	1	1	1	1		1	2	2	1
	Hume and Odom (2007)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	2	2	1

Table D5

IRR Percent Agreement for Each Quality Indicator

Basic Design Standards						
	Systematic Manipulation	**IOA Recorded**	**IOA Session Totals**	**IOA Percent Agreement**	**Attempts to Present Effect**	**Data Points**
Percent Agreement	100%	95%	82%	87%	100%	95%
Descriptive Standards						
	Participant Description	**Setting Description**	**Interventionist Description**	**Baseline and intervention Procedure Description**	**Dependent Variable Description**	
Percent Agreement	74%	44%	67%	67%	74%	
Additional Measures Quality Standards						
	Maintenance	**Generalization**	**Fidelity**		**Social Validity**	
Percent Agreement	72%	77%	69%		67%	
Evidence Standards						
	Baseline Analyses (baseline change, predictability, consistency of variability, and trend)	**Within Phase Analysis** (data points, predictability, consistency of variability, and trend)	**Between Phase Effects** (basic effect, immediacy change in level, immediacy change in trend, overall change in level, overall change in variability, and overlap, and similar phase data pattern)		**Overall Effectiveness** (data points, treatment effect, effect ratio, and evidence of effect)	
Percent Agreement	88%	89%	94%		96%	

APPENDIX E

POV OUTCOMES AND SURVEYS

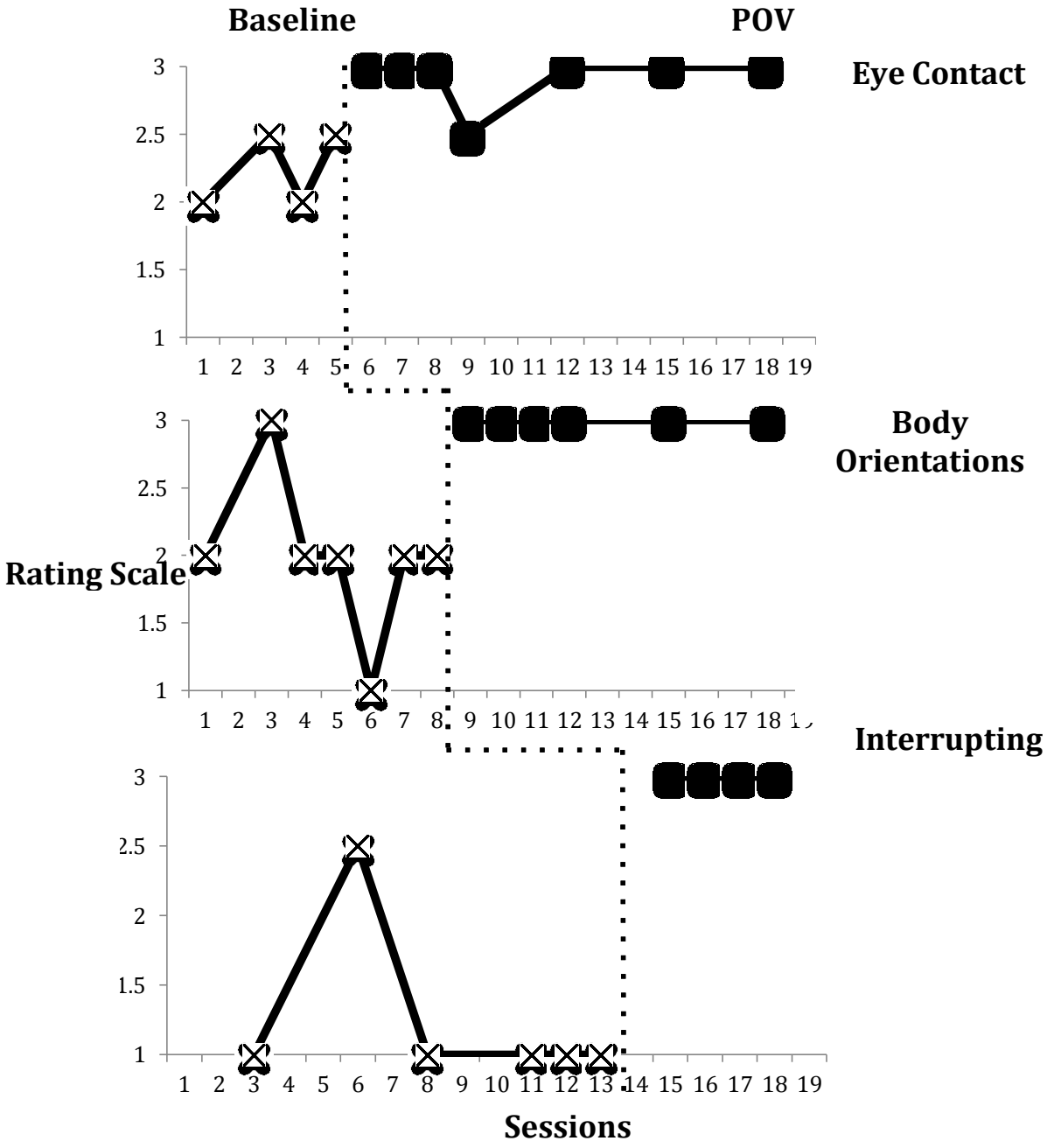

Figure E1. John's multiple-baseline data across three target conversational skills.

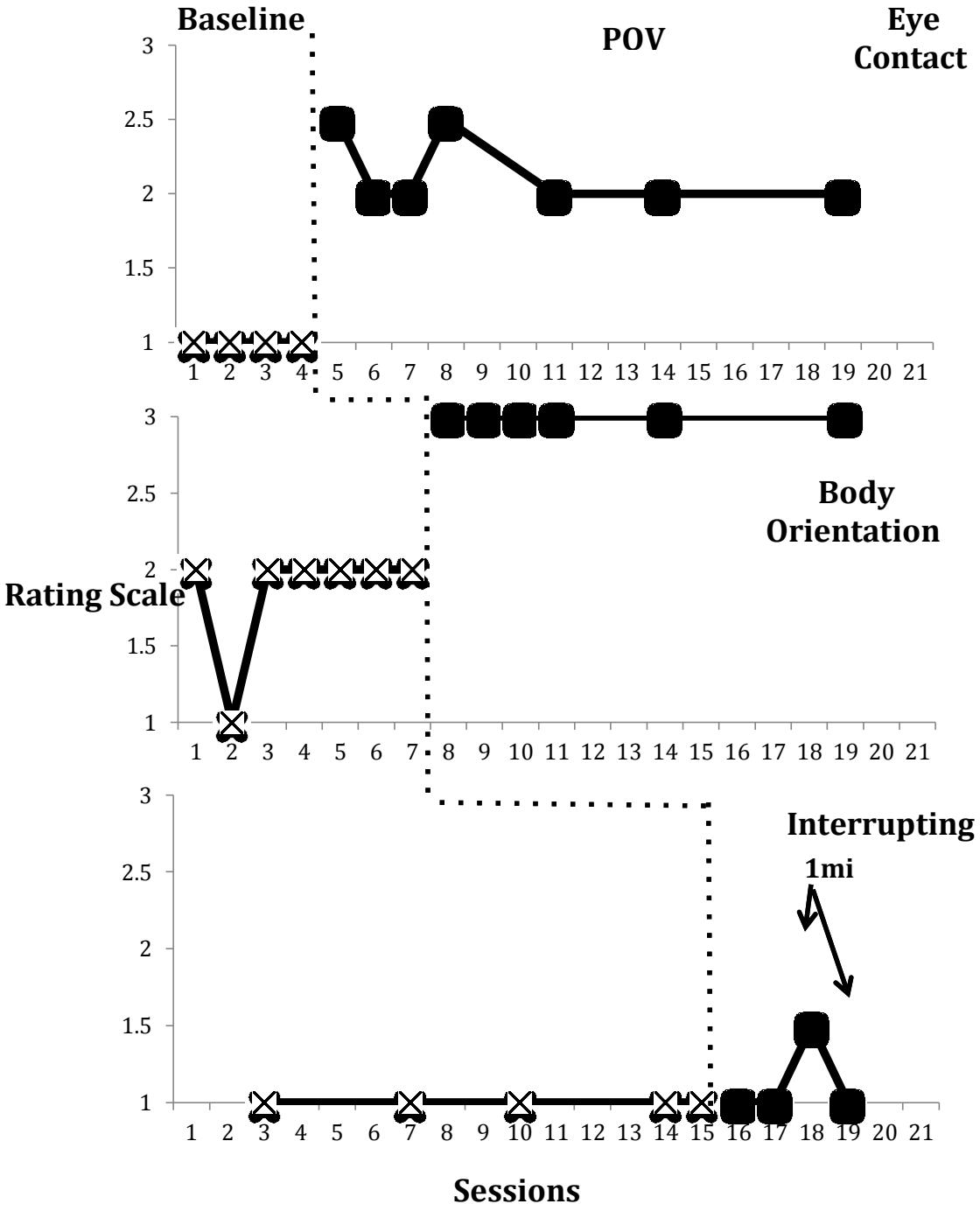

Figure E2. Steven's multiple-baseline data across three target conversational skills.

Table E1

Skill Rating Measures

Eye Contact				
1	**1.5**	**2**	**2.5**	**3**
No eye contact or very little was given (constant eye shifting or staring into the distance)	eye contact occurs, but it is inapprpriate (more than 5 seconds of eye contact)	Appropriateeye contact happens, but it happens rarely, or there is both appropriate and inappropriate eye contact	Appropriate Eye Contact happens (3-5 seconds and then shifts eyes) for about half of the conversation	Appropriate eye contact happens (3-5 seconds and then shifts eyes) for the majority of the conversation
Body Orientation				
1		**2**		**3**
The body (shoulders, knees, and head) are all NOT facing the person talking		One or two portions of the body (shoulders, knees, or head) are facing the speaker		The entire body (shoulders, knees, and head) is facing the speaker
Interruption				
1	**1.5**	**2**	**2.5**	**3**
Interrupts with two to three word sentences more than once	Interrupts with a sound or partial word verbalization more than once	Interrrupts with two to three word sentences only once	Interrupts with a sound or partial word verbalization more than once	No interruptions

Table E2

Pre-Study Survey

#	Survey Questions	Answer Options		
1	Do you feel comfortable when talking to a group of people you are friends with?	No	I don't Know	Yes
2	Do you feel or would you feel comfortable when you are working and have to talk to customers?	No	I don't Know	Yes
3	Do you feel comfortable talking to your teacher?	No	I don't Know	Yes
4	Do you feel nervous when you are around others?	No	I don't Know	Yes
5	Do you feel scared when you are around others?	No	I don't Know	Yes
6	Do you think it is easy to talk to others?	No	I don't Know	Yes
7	Do you think it is difficult to talk to others?	No	I don't Know	Yes
8	Is it difficult to make eye contact with others?	No	I don't Know	Yes

Table E3

Social Validity Survey

#	Social Validity Questions	Answer Options		
1	Did you like technology tools we gave you?	Yes	Sometimes	No
2	Do you think the technology tools we gave you were helpful?	Yes	Sometimes	No
3	Would you like the videos to be on your own iPod or iPhone?	Yes	Some of the Videos	No
4	Would the videos be helpful if you watched them before you went into a situation when you had to talk with other people?	Yes	Sometimes	No
5	Did you like the device that we gave you that recorded your voice and what you did?	Yes	Sometimes	No
6	Do you feel like we were too intrusive/we got in your way too much?	Yes	Sometimes	No