Research and Practice in the Schools

TASP Texas Association of School Psychologists
A partner on the path to student success

Research and Practice in the Schools:
The Official Journal of the Texas Association of School Psychologists

_____

Research and Practice in the Schools:
The Official Journal of the Texas Association of School Psychologists

Research and Practice in the Schools is a publication of the Texas Association of School Psychologists (TASP). It is an online, peer-reviewed journal that provides TASP members with access to current research that impacts the practice of school psychology. The primary purpose of *Research and Practice in the Schools* is to meet the needs of TASP members for information on research-based practices in the field of school psychology. To meet this need, the journal welcomes timely and original empirical research, theoretical or conceptual articles, test reviews, book reviews, and software reviews. Qualitative and case-study research designs will be considered as appropriate, in addition to more traditional quantitative designs. All submissions should clearly articulate implications for the practice of psychology in the schools.

Instructions for Authors

General Submission Guidelines

All manuscripts should be submitted in electronic form to either of the co-editors (jeremy.sullivan@utsa.edu or art.hernandez@tamucc.edu) as an email attachment. Manuscripts should be submitted in MS Word format and labeled with the manuscript's title.

It is assumed that any manuscript submitted for review is not being considered concurrently by another journal. Each submission must be accompanied by a statement that it has not been simultaneously submitted for publication elsewhere, and has not been previously published.

Authors are responsible for obtaining permission to reproduce copyrighted material from other sources. IRB approval should have been obtained and should be noted in all studies involving human subjects. Manuscripts and accompanying materials become the property of the publisher. Upon acceptance for publication, authors will be asked to sign a publication agreement granting TASP permission to publish the manuscript. The editors reserve the right to edit the manuscript as necessary for publication if accepted.

Submissions should be typed, double-spaced with margins of one inch. All articles should meet the requirements of the *APA Publication Manual, 6th ed*., in terms of style, references, and citations. Pages should be numbered consecutively throughout the document. Illustrations should be provided as clean digital files in .pdf format with a resolution of 300 dpi or higher. Tables and figures may be embedded in the text. A short descriptive title should appear above each table with a clear legend and any footnotes below.

The Review Process

After receiving the original manuscript, it will be reviewed by the Editors and anonymously by two or more reviewers from the Editorial Board or individuals appointed on an *ad hoc* basis. Reviewers will judge manuscripts according to a specified set of criteria, based on the type of submission. Upon completion of the initial review process, feedback will be offered to the original (primary) author with either (a) a preliminary target date for publication; (b) a request for minor editing or changes and resubmission; (c) significant changes with an invitation for resubmission once these changes are made; or, (d) a decision that the submission does not meet the requirements of *Research and Practice in the Schools*.

Article

Utilizing Video-Self Modeling for Adolescents with Intellectual Disabilities in the Educational Setting

Whitney Appleby, Wendi L. Johnson, LaToya Bowens,
Krystal Franco, and Brynn Reece
Texas Woman's University

Video self-modeling (VSM) is an intervention in which a student views a video of him/herself correctly performing a target behavior or skill in order to improve the performance of the skill. This article will detail a research-to-practice intervention pilot study in which middle school and high school students with intellectual disabilities participated in an intervention to improve their functional and social skills. A pre-post-test design was employed across 8 secondary students eligible for special education services due to an intellectual disability. Individualized goals were selected after interviews were completed with the student, teacher, and parent/guardian, as well as baseline observations and standardized data collection. Post-intervention data were collected in the form of parent and teacher feedback, and standardized data to determine the potential change in the targeted behavior. Results from the observational data and interviews demonstrated positive change in the targeted behaviors, and a reduction in the negative impact of the behavior on the students' educational progress. The standardized measures did not demonstrate sensitivity to the small positive change noted. Nonetheless, VSM within an educational program appears a promising intervention for supporting the development of social and functional skills for adolescents with intellectual disabilities. Findings from this study support a call to future scholarship in the use of VSM with this population.

Keywords: Video Self-Modeling, intellectual disability, social skills, functional skills

Modeling and observational learning are taking place intentionally and unintentionally on an everyday basis. By observing a behavior being performed by another, the learner can imitate and then integrate a new skill. According to Bandura (1986), modeling serves as a social prompt to promote similar behaviors and actions in others. It has been shown to be both an effective teaching tool for children and adolescents, as well as an essential component of their learning experience (Buggey & Ogle, 2012; Prater, Carter, Hitchcock, & Dowrick, 2012). While adult and peer models are critical to a child's development and learning, it is believed that the self-as-a-model approach offers a unique and enhanced opportunity for one to successfully learn a new skill, improve upon an old skill, as well as increase self-efficacy for engaging in the behavior (Bandura, 1986; Dowrick, 1999).

Video Self-Modeling

Video self-modeling (VSM) is a form of the self-as-a-model approach that is currently being used in research and practice. The theoretical foundation for VSM is based on Bandura's social learning theory (1977, 2001), which proposes that the observer learns best when the model shares similar characteristics to the viewer such as age, physical characteristics, and skill level. This aspect of social learning is also enhanced when the ob-

Author's Note: Correspondence concerning this paper should be addressed to Wendi L. Johnson, Department of Psychology and Philosophy, Texas Woman's University, P.O. Box 425470, Denton, Texas 76204. E-Mail: wjohnson4@twu.edu

server believes that the target goal is attainable. Miklich and Creer (1974) proposed that having an individual observe him/herself performing a skill on videotape increased the likelihood to attend to the video, and thus increased the chances of learning the new skill.

The most common forms of VSM used for children are the feed-forward method and positive self-review. The feed-forward method is used when images are provided of the child performing developmentally appropriate behaviors that they have not performed before or within that behavioral sequence. For example, if a child has previously only exhibited the use of single-word speech, a feed-forward VSM intervention movie may demonstrate the child speaking a full sentence by splicing together the single words obtained during recording into a cohesive sentence. With the positive self-review method, children view images of behaviors that are currently in their repertoire of skills but occur in low frequency (Buggey, 2009). In both methods, the children observe themselves on video performing positive, adaptive behaviors (Dowrick, Kim-Rupnow, & Power, 2006). In order to accomplish this task, video footage is collected in a variety of ways, such as prompting the children to say rarely used words or phrases, having the children engage in challenging behavior through role play, or letting the camera roll to capture rare behaviors (Buggey, 2009). The videotape is shown to the children only after the video footage has been edited to show an errorless performance of them engaging in the desired behavior or skill (Prater et al., 2012).

Application of Video Self-Modeling

VSM has been shown to increase positive behaviors in children and adolescents across varying ages, disabilities, and skills, as well as across diverse settings (Buggey & Ogle, 2012). Research also supports the effectiveness of VSM when teaching students social and functional skills. For example, VSM has been successful at increasing and helping to maintain social initiation (Buggey, 2005; Buggey, Hoomes, Williams, & Sherberger, 2011), improving social interactions such as turn-taking and social questions (Lantz, 2005), increasing peer social engagement in a natural setting (Bellini, Akullian, & Hopf, 2007), and in cultivating emotional awareness (Bernard-Ripol, 2007) in children ranging in age from 4 to 18.

Functional skills have also been increased for individuals using VSM. Studies have shown that VSM is effective for increasing adaptive tasks, such as bed-making skills (Miklich, Chida, & Danker-Brown, 1977), and complex behavior modification, such as increasing worker production with adult participants (Dowrick & Hood, 1981). Furthermore, VSM has been shown to increase task fluency for a variety of self-help tasks necessary for independent living (Lasater & Brady, 1995) and work skills, such as simple office tasks (Cihak & Schrader, 2009).

Although VSM research has involved children and adolescents with a range of disabilities, the primary focus of research on this intervention has involved children with autism spectrum disorder (Bellini & Akullian, 2007). Little research has evaluated the efficacy of VSM with children with intellectual disabilities (ID). Children with ID often display deficits in their social and functional skills that impede their performance in school and their adaptive functioning in society, suggesting this population may benefit from an intervention targeting these areas. The limited research that has sought to determine the efficacy of VSM with children with ID has focused primarily on behavioral goals. In one of the first research studies using self-modeling for a child with ID, Brown and Middleton (1993) found that the implementation of a self-modeling intervention produced a dramatic decrease in self-stimulating behavior. Expanding upon these research findings with this population, Bilias-Lolis, Chafouleas, Kehle, and Bray (2012) employed a multiple baseline design with 3 high school students with ID evaluating the efficacy of a video self-modeling intervention on the reduction of disruptive classroom behavior. Each student demonstrated a decline in disruptive behavior as measured by visual analysis which was further supported with the calculation of large effect sizes ranging from 1.86 to 2.27. While both studies demonstrate the promise of VSM as beneficial for the reduction of self-stimulating or disruptive behavior, additional research is necessary for VSM as an intervention for social and functional skill

development in children with ID. The purpose of this pilot study was to further extend the literature on VSM as it relates to social and functional skill attainment within a school-based setting for secondary students with ID. This study also provides a larger sample size than previously provided for similar populations, as previous studies have typically included one to five participants [e.g., Bellini, Akullian, & Hopf, 2007 (two students with autism for social development); Bernad-Ripol, 2007 (one student with Asperger Syndrome for emotional identification); Bilias-Lolis, Chafouleas, Kehle, & Bray, 2012 (three students with ID for disruptive behavior); Buggey, 2005 (five students with autism for social engagement)].

Methods

Participants and Setting

Eight secondary students ranging in age from 12 to 21 years participated in this pilot study. These students were recruited from a suburban public school district in North Texas. The participants consisted of 5 females and 3 males, and the ethnicities represented were Caucasian/ Non-Hispanic, Hispanic, and African American. All of the participants were eligible for special education services due to an intellectual disability. The range in intelligence quotient (IQ) of the participants varied from 19 to 77 on a standardized cognitive assessment measure. Due to this eligibility, all participants required placement in a self-contained classroom where their individualized education plans (IEP) could be implemented which required substantial modifications and accommodations to the standard curriculum for each grade level. As such, these students' academic goals were written below grade level and often focused on prerequisite or foundational skills. For example, functional life-skills goals focused on self-care and hygiene tasks, building and developing social skills, as well as adaptive skills such as communication and community use.

Measures

The special education teachers and the participants' parents were given a pre- and post-intervention interview that elicited information about the impact of the student's target behavior in the educational setting. Seven Likert-scale type questions were utilized as part of the interview to examine how much the target behavior impacted the student's educational performance by responding to questions such as, "On a scale of 1-10, how much does this behavior negatively impact the student's educational progress?" The score of 1 indicates that the behavior does not significantly impact the student's educational progress negatively. The score of 10 indicated that the behavior significantly impacted and hindered the student's educational progress. The pre-intervention information was compared to the post-intervention information given by the teacher. Additionally, the teacher's rating of how much the behavior has positively changed or improved within the educational setting was examined. Pre-intervention and post-intervention information from the parents of the participants was compared to survey the extent to which improvement was evident and generalized outside of the classroom environment.

Two pre-intervention observations, as well as two post-intervention observations, were completed. The interventionists received extensive training on how to collect observation data, and each observation was completed across a 30 minute period in the participant's primary special education classroom. Observational data were recorded via a frequency count within one minute intervals. To maintain consistency across the observations, the same interventionist observed each participant and recorded data on the same target behavior across all four observations. The observation time and location within the classroom were also stable whenever possible based on the student's educational programming. This method was replicated during post data collection to evaluate the potential change in the frequency of the target behavior in the classroom environment.

The Adaptive Behavior Assessment System-II (ABAS-II; Harrison & Oakland, 2003) was also utilized with the parent and the teacher during the pre- and post-intervention periods. The ABAS-II is a comprehensive, norm-referenced rating scale to measure skill areas such as communication, self-direction, social, functional academics, motor, self-care, leisure activities, as well as health and safety

choices. The respondent completes the ABAS-II by responding if the individual is able to complete the task independently and if so at what frequency (i.e., *Always, Sometimes,* or *Never*). Both the parent and teacher completed this measure to determine the consistency and frequency of the targeted behavior across settings. The Behavior Assessment System for Children – Second Edition (BASC-2; Reynolds & Kamphaus, 2004) parent and teacher forms were also utilized as a standardized measure for baseline assessment and post-intervention assessment of skill level. The BASC-2 is a behavior rating scale that measures a wide variety of behaviors and adaptive skills. Each item on the BASC-2 is rated using a four-point Likert-scale (i.e., *Never, Sometimes, Often, Almost Always*). For the purposes of this study, only the Adaptive Behavior composite scores were utilized. The composite score is a combination of the adaptive scales which examine the participant's ability in the areas of Adaptability, Social Skills, Leadership, Activities of Daily Living, and Functional Communication. The ABAS-II and BASC-2 were selected for the purpose of evaluating the level of change in the adaptive skills of the participants using standardized measures based on both parent and teacher reports. It was hypothesized that these standardized measures would show improvement in adaptive skill areas after the participant received the VSM intervention targeting the social skill or functional behavior.

Procedures

All procedures utilized were approved by the university Institutional Review Board to ensure ethical practice and protection of all participants. For the current pilot study, consent was obtained from the parents and the teachers, as well as assent from the minor students. After attaining consent, the intervention proceeded across three phases. During Phase One, the baseline assessment data were obtained on the target behavior by gathering information from the teacher and parent through a pre-intervention interview, as well as through the observation data. The ABAS-II and BASC-II were also completed by both the parent and the teacher. Each participant had a goal to improve a functional life skill or a social skill area

that supported the student's IEP. The target goal was individualized to each participant and agreed upon by the teacher and parent. Phase One was completed during the early fall of the school year.

Phase Two consisted of six small group sessions which began with building rapport with the students. The middle school and high school students were grouped together respectively. The second session consisted of having the students practice role-playing the desired, targeted behavior, as practicing strengthened their skills and helped the students become familiar with the video camera in the room. During the third and fourth sessions, video-taping occurred while students were provided with various levels of support to engage in the desired behaviors. For example, one student who was working on utilizing her communication device acted out various "real-life" scenarios such as using her device at work with her boss. Support was provided during the role-play situations by the interventionist through verbal directions, physical modeling, or gestural cues to obtain the desired behavior on video. Once several clips of the positive targeted behavior were captured on video, the verbal or gestural prompts were eliminated from the raw footage. The final product was a 2-3 minute individualized "movie" in which the participant was presented in several short clips engaging in the preferred behavior and being successful. The fifth session of Phase Two began with having the students watch their movie. To ensure integrity of the intervention, the teacher completed a daily chart documenting that the student watched their movie one time per day for 10 consecutive school days. Treatment integrity was supervised through frequent contact with the teacher via email or phone call to monitor how the student responded to the video and if there were any changes in the routine of the presentation. The viewing presentation of 10 consecutive school days was chosen based on research that demonstrates that if changes are to occur, they will take place within a short time after viewing (Buggey, 2007; Dowrick, 1983). The sixth and final session of Phase Two concluded by following up with the participants to discuss how they felt about their movie and talk about the successful behavior that was evidenced. Generalization of the skill was discussed with the

parents and the teachers, as well as maintenance of the gains that were reported.

Small levels of reinforcement were also established concurrently with the intervention. Verbal praise was abundantly provided during the practice session 2, as well as during the taping sessions, to encourage the participant toward the desired behavior. Additionally, during the video-viewing time each participant was provided with a small daily reinforcer (i.e., a sticker or smiley face on their chart). This allowed for immediate reinforcement for watching their movie. These small reinforcers were chosen based on the previously established reinforcement system utilized within the student's class. The participants were already familiar with this type of reinforcement and its effectiveness had been verified in the classroom prior to the intervention. Additionally, within each participant's movie there were small phrases such as "Great job Maggie" or auditory reinforcement such as clapping or music. At the end of the VSM intervention, a celebration party was provided for each participant with their class. These reinforcers as part of the VSM intervention were consistent with the already established classroom reinforcement that was in place prior to the intervention implementation.

At the conclusion of the six session sequence, Phase Three began. During this phase, data were obtained by gathering information from the teacher and parent through a post-intervention interview and two follow-up observations were conducted in the classroom of the identified target behavior for each student. Consistency was maintained, as the same interventionist who completed the pre-intervention observations and interventions also completed the post-intervention data collection. Parents and teachers were provided with the results of the pre- and post- intervention data in graphed form, and parents were provided a copy of the VSM movie if desired. The ABAS-II and BASC-II were also completed by both the parent and the teacher after the intervention to determine a possible change in the targeted behavior. Phase Three was completed by the end of the student's school year. Following the completion of all three phases of the VSM intervention, the parents of the participants were provided with a personal copy of their child's VSM video. In addition, with parental consent the participants' teachers were also provided with a copy of the video for future use.

Data Analysis

Data were analyzed by evaluating teacher responses to the pre- and post-intervention interview questions. Teachers were asked to rate how much on a 1 to 10 Likert-type scale the student's target behavior negatively impacted their educational performance. Using this rating for each student before and after the implementation of the intervention, percentages were calculated to measure how much the intervention package assisted in minimizing the impact of the skill deficit on the child's educational performance. A positive change in the rating indicated less of an impact of that behavior or skill deficit on the child's educational performance after the intervention. Additionally, teacher ratings of the percentage of positive change in the target skill or goal for each student were evaluated. Narrative feedback was also examined from the pre- and post-intervention parent interview. Data from the pre- and post-intervention direct observations of the participants were reviewed. Additionally, mean percentages were calculated for the occurrence of the target behavior or skill during baseline collection and post-intervention. These percentages were compared to determine the amount and direction of any change in the target skill after the implementation of the intervention.

Standardized data from the ABAS-II and BASC-II were also examined to determine the possible evidence of positive change in the targeted behavior. Specific subscales that focused on the student's targeted goal were chosen for comparison. For example, for a student whose goal was to increase her ability to initiate conversations, the results from the Social subscale of the ABAS-II were examined. For the BASC-II results, only the scores from the Adaptive Behavior composite were reviewed given the focus on this study. A Reliable Change Index (RCI) score was calculated for each participant on both standardized measures. The RCI was first proposed by Jacobson and Truax (1991) as a measurement of change for an individual from baseline to post-treatment. This

allows the researcher to examine a participant's response to an intervention by comparing scores on an outcome measure. The RCI utilizes the standard error of measure (SEM) to determine if improvement demonstrated from baseline to post-treatment is due to chance variation or a true change in behavior (Schmitt et al., 2013). The RCI is calculated by subtracting the participant's post-intervention score by the baseline score and then dividing it by the standard error of the difference, which is based on the SEM of the assessment instrument. The RCI scores were calculated on selected subscales of the ABAS-II and the BASC-II adaptive behavior composite score for each participant.

Results

From the data collected through the interviews, of the eight students who participated in this pilot study, seven of the teachers described the target behavior as having less impact on the participant's educational progress after the intervention. The teachers reported a mean percentage of decrease as 48% from pre-intervention to post-intervention. Additionally, teachers' perceptions of positive change for all students in their targeted behavior or skill averaged 50%.

Observational data for the students' target behaviors were available for four out of the eight participants. Four students did not have observational data as a result of the target goal being a low frequency behavior and the behavior was not demonstrated during either baseline or follow-up observations. Of the four students for which observational data were elicited, Student 1 showed an increase in maintenance of eye contact during social introductions with a mean frequency of 25% during baseline to 50% at follow-up. Student 2 had a goal to increase the frequency of her independent use of an assistive technology device to communicate with others during unstructured time. Across the pre- and post-intervention observations she showed a positive increase in the frequency of her device with a mean of 15% usage at baseline and a mean of 52% at follow-up. Student 3 had a goal of increasing her frequency of polite responses, which included the

use of positive vocal tone, eye contact, and smiling to statements made by teachers and classmates. Student 3 demonstrated these behaviors with a mean of 37.5% at baseline to 47% at the follow-up observations. Lastly, Student 4 increased the frequency of responding within a five second interval to an adult's question or request from a mean of 55% at baseline to 70% at follow-up. Overall, each set of observational data showed improvement across the participant's individual skills, with some students demonstrating greater gains than others. Specifically, overall percentage of positive change for each of the four students was 25%, 37%, 9.5%, and 15%, respectively across the pre- and post-intervention observations.

Despite observational data of low frequency behavior not being available for the remaining four participants, feedback from the teacher ratings during the interview demonstrated improvement. For example, one of the students demonstrated significant improvement in appropriately identifying her feelings during "meltdowns" utilizing her Picture Exchange Communication System (PECS). Meltdowns were defined as crying, screaming, and non-compliant verbal and physical refusal. The PECS system was utilized in the classroom and during the VSM intervention to assist the student in verbalizing her emotions when she became upset. The teacher shared the following information during the follow-up interview:

> During [her] most recent meltdown, which occurred after the VSM intervention, I noticed that [she] was trying to tell us her feelings by saying, "Mad, mad, mad," which is something she did not do before. The meltdown was also less severe, and we were able to acknowledge her feelings.

During the teacher interview, this student was reported as demonstrating a 40% positive change in her target behavior. Additionally, for the other three participants without observational frequency data, the teacher's rating of positive change was examined. The three remaining students were rated as having a 10%, 50%, and 70% positive change in their target social or functional goal from the teachers' perspective.

Despite multiple attempts to obtain parental feedback post-intervention, responses were acquired from five parents of the eight participants. During

the parent interview, the parents provided feedback regarding the impact of their child's target behavior on the educational process. Of these five parents, two described no change post-intervention while three parents reported decreases of 20%, 30%, and 30% of the negative impact of the child's target behavior on his or her educational progress. All eight participants had parent feedback regarding their perception of positive change post-intervention. Across all students, an average of 35% change in the positive direction was reported by each participant's parents.

For each participant, RCI scores were calculated for the BASC-II Adaptive Behavior composite scores and the ABAS-II subscale scores from both the parent and teacher rating scales of each measure. Analysis of the RCI calculated for the BASC-II resulted in non-significant scores. The individualized subscale scores from the ABAS-II, which corresponded with each participant's target goal, were also non-significant.

Discussion

The primary purpose of this research-to-practice pilot study was to examine the efficacy of VSM when implemented concurrently with classroom interventions in a public school special education setting. Participants were middle school and high school students with social and/or functional deficits who each met special education eligibility as a student with an intellectual disability.

According to the parent and teacher ratings obtained during the interviews, the VSM intervention led to an increase of the targeted positive behaviors for seven of the eight participants, which assessed each teacher's perception of how much the target behavior impacted the student's academic performance. It should be noted that the one student for whom the teacher did not report improvement did demonstrate progress in their data collected through pre- and post-intervention observations. The observations documented the number of instances where the appropriate replacement behavior was utilized before and after VSM was implemented with an improvement of 37%. It should also be noted that four participants' parents reported that they were

unsure of the degree of negative influence the behavior was having in the classroom setting.

Based on frequency counts for the observational data, VSM interventions led to varying degrees of improvement by increasing the amount of replacement behaviors or skills. As noted earlier, only four of the eight participants had available observational data due to the target goal being a low frequency behavior and therefore not observed during the pre- and post-intervention observations. However, all four students with observational data improved toward their social or functional goal after the VSM intervention.

Due to the nature of the study design, the researchers cannot say with certainty whether or not the intervention was efficacious. Despite this fact, the findings were noteworthy in that each participant demonstrated an increase toward their goal either through observational data, pre- and post-intervention interview data provided by the teacher, or both. Although less robust findings were evident, pre- and post-intervention data provided by parents suggested that many of the students' increase in target skills were generalizable to settings outside of the special education classroom setting. In addition, the intervention was a feasible method for a school-based intervention and was largely well received by school personnel, the participants, and their families. The investigators posit this study provides preliminary evidence for the efficacy of VSM for children with intellectual disabilities. Findings from this pilot study are promising and warrant a call for future scholarship in this area to unconditionally recommend the intervention for children with intellectual disabilities.

Limitations and Future Research

Limitations of this research-to-practice pilot study and the conclusions drawn within this article need to be addressed due to the nature of the research design and sample size. Interview data, which are useful in understanding the target behavior, were based on the parents' and the teachers' perception of the target behaviors and are therefore subject to personal biases. The respondents may have wanted the participant's

behavior to demonstrate improvement and therefore rated the outcome in a more positive manner.

Additionally, statistical analysis is limited due to a sample size of eight participants. While this *n* is larger than many comparable studies, it is often difficult to determine significance with traditional statistical analyses with such a small sample. In addition, the sample was further reduced for observational data (*n*=4) and parent responses (*n*=5). Although qualitative findings from this study suggest that the intervention was successful, these findings were in contrast to the non-significant findings found in the current sample using the RCI statistic. The RCI statistic does allow for small *n* calculations to determine the possibility of pre- and post-intervention change for an individual; therefore, the contrast in qualitative and quantitative findings is hypothesized to be an effect of the standardized measures utilized. While both the BASC-II and the ABAS-II have very strong reliability statistics, their domain scores provide a measure of performance or functioning on a broad scale (i.e., social, functional academic, etc.). Because the individualized goals for each participant were specific in nature, such as addressing a single social skill, the broad subscale measures on the BASC-II and ABAS-II were likely not sensitive enough to demonstrate change in the specific domain areas. Additionally, the norm-referenced measures were utilized in the early fall and the late spring during one school year so the length of time between the pre- and post-intervention evaluation may not have been long enough to detect change on these global measures. Furthermore, the investigators acknowledge that the use of percentages as a unit of measurement serves as a limitation as it does not take into account the stability of the measures. However, the percentages were included to provide the reader with a quantifiable way of understanding the extent to which change was observed based on the feedback provided from the parents and teachers. As such, significance of the level of change cannot be determined and these findings should be interpreted with caution.

While attempts were made to minimize observer bias in this pilot study by providing extensive training and maintaining consistency across observers, location, and time of the observations for each participant, it is still possible that observer bias could have had an effect on the results. The observers likely desired positive outcomes due to the time invested in the intervention which could have possibly inadvertently skewed the observational data. The investigators and observers were not blind to the participants of the study, as the same observer who completed pre- and post-observations also worked with the same participants during the intervention phase. Future research should include additional observations and integrate inter-observer reliability ratings to account for this potential bias. Additionally, a control group could be utilized in future research to ensure that the observers for the pre and post data are unaware of whether or not a participant did or did not receive the intervention. Furthermore, the decision to target low frequency behaviors was also a challenge in this pilot study. While agreement was obtained between the teachers and the parents regarding the target goal after extensive discussions, future researchers should complete preliminary observations to determine if the chosen behavior is demonstrated within that setting at a level amenable to this type of intervention.

The functioning level of the adolescent may have also impacted the outcomes. While all of the participants were in a similar self-contained educational setting based on their special education eligibility of intellectual impairment, each individual demonstrated a wide variability of skills. For example, in the area of language functioning one student was non-verbal while another student needed to work on pausing before responding to a verbal request. The assessed IQ level of each participant was also a broad range. It is hypothesized that the most significant outlier of IQ levels (a SS of 19) was primarily due to the verbal loading on the measure given to determine intellectual functioning. The student was able to communicate through non-verbal avenues and gestural cueing; however, this was not well represented with the standardized cognitive measure and likely resulted in an IQ score much lower than the student's actual ability level. Another possible confound to the results relates to the other interventions that were taking place concurrently with the VSM intervention. Within each classroom

setting the teacher had in place a reinforcement system prior to the VSM intervention for the students. For example, in one classroom there was a monetary system set up where the students could earn pretend money when complying with classroom expectations and then spend it at the "store" at the end of the week for small trinkets and prizes. Additionally, several of the students had behavioral intervention plans in place as part of their IEP which delineated specific recommendations for behavioral issues within the classroom. All of these systems were running concurrently to the VSM intervention and may have increased the likelihood of finding a positive effect. Also, the process of going through the intervention itself could have impacted the results. Buggey (2009) discusses how the act of recording the student's behavior often elicits positive behavior. The additional scaffolding supports, which were provided through the videotaping process by practicing the target behavior and acting out the various role-play scenarios, also likely increased the chances of a positive outcome with the VSM intervention. Future research may want to examine this phenomenon and determine if the positive outcomes reported across studies utilizing VSM are a result of the process of the VSM intervention itself or possibly the combined effect of multiple concurrent interventions.

The location of the intervention may have also had an effect. The participants were split between two different schools with different teachers. This may have further impacted the data collected through teacher responses on pre- and post-interviews. Also, post-intervention data were collected after only 10 viewings of the VSM videos with an average of eight weeks between pre- and post-data collection periods. While significant gains were apparent after the two week intervention implementation period, more improvement may be observed after an even longer implementation period. The short data collection period prohibited gathering information on the long-term effects and maintenance of VSM. More research is needed in this area to help continue building the foundation of evidence supporting the use of VSM with diverse populations. Also, current VSM research varies in opinion regarding the most effective setting (e.g., school vs. home), implementation and viewing time

period, and other details such as the use of the spacing effect (i.e., watching the movie every other day or with multiple times in between viewing times vs. every day for a set period of time), maintenance of new skills over time, and the type of VSM utilized (i.e., positive self-review vs. feed forward). Future research should address these areas to increase the overall utility of VSM and to know which aspects of VSM are most efficacious for specific populations.

Summary

This pilot study provided an extension of the current literature base by demonstrating the application of VSM as a promising intervention for an adolescent population with intellectual disabilities. In addition, this article demonstrates the need for additional research to better verify the use of the intervention as an evidence-based practice for this population. The number of participants (i.e., 8 secondary students) in this pilot study is an increase in comparison to similar other VSM published studies, which have an average sample size of one to five participants (Bellini, Akullian, & Hopf, 2007; Buggey, 2005). The results of this study add to an emerging body of research demonstrating the efficacy of VSM in adolescents with social and functional goals. Despite the standardized measures not demonstrating sensitivity to the specific goal attainment for each student, data from the parent and teacher interviews as well as observational data demonstrated improvement on behavioral, social, or functional goals after watching their individualized VSM video 10 times over a two week implementation period. While the majority of the VSM literature focuses on individuals with autism, this intervention modality is unique in that it can be individualized based on the participant's functioning level and can be used to augment educational goals that are already in place as part of the student's educational plan. VSM allows for the participants to view themselves being successful in an area of past difficulty, which as a result can increase self-confidence in that skill area and potentially lead to increased generalization across settings. These distinctive features of VSM make it an excellent fit for being utilized in the school environment. As demonstrated in this pilot study,

VSM can support social, functional, or behavioral goal attainment to supplement a student's IEP and potentially improve overall functioning in the educational setting.

References

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.

Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology, 52,* 1-26. doi: 10.1146/annurev.psych.52.1.1

Bellini, S., & Akullian, J. (2007). A meta-analysis of video modeling and video self-modeling interventions for children and adolescents with autism spectrum disorders. *Exceptional Children, 73,* 264–287. doi: 10.1177/001440290707300301

Bellini, S., Akullian, J., & Hopf, A. (2007). Increasing social engagement in young children with autism spectrum disorders using video self-modeling. *School Psychology Review, 36,* 80-90.

Bernad-Ripol, S. (2007). Using a self-as-model video combined with social stories to help a child with Asperger syndrome understand emotions. *Focus on Autism and Other Developmental Disabilities, 22,* 100–106. doi: 10.1177/10883576070220020101

Bilias-Lolis, E., Chafouleas, S. M., Kehle, T. J., and Bray, M. A. (2012). Exploring the utility of self-modeling in decreasing disruptive behavior in students with intellectual disability. *Psychology in the Schools, 49,* 82-92. doi: 10.1002/pits.20616

Brown, G., & Middleton, H. (1998). Use of self-as-a-model to promote generalization and reduction of self-stimulation in a child with mental retardation. *Education and Training in Mental Retardation, 33*(1), 76-80.

Buggey, T. (2005). Applications of video self-modeling with children with autism in a small private school. *Focus on Autism and Other Developmental Disabilities, 20,* 180–204.

Buggey, T. (2007). A picture is worth: Video self-modeling applications at school and home. *Journal of Positive Behavior Interventions, 9,* 151-158.

Buggey, T. (2009). *Seeing is believing*. Bethesda, MD: Woodbine House.

Buggey, T., Hoomes, G., Sherberger, M. E., &Williams, S. (2011). Facilitating social initiations of preschoolers with autism spectrum disorders using video self-modeling. *Focus on Autism and Other Developmental Disabilities, 26*(1), 25–36. doi: 10.1177/ 1088357609344430

Buggey, T., & Ogle, L. (2012). Video self-modeling. *Psychology in the Schools*, *49*, 52-70. doi: 10.1002/pits.20618

Cihak, D., & Schrader, L. (2009). Does the model matter? Comparing video self-modeling and video adult modeling for task acquisition and maintenance by adolescents with autism spectrum disorders. *Journal of Special Education Technology, 23*, 9–20.

Dowrick, P. W. (1983). Self-modeling. In P. W. Dowrick & J. Biggs (Eds.), *Using video: Psychological and social applications* (pp. 105-124). New York: Wiley.

Dowrick, P. W. (1999). A review of self- modeling and related interventions. *Applied & Preventive Psychology*, *8*, 23-39. doi: 10.1016/S0962-1849(99)80009-2

Dowrick, P.W., & Hood, M. (1981). Comparison of self-modeling and small cash incentives in a sheltered workshop. *Journal of Applied Psychology, 66*, 394–397. doi: 10.1037/0021-9010.66.3.394

Dowrick, P. W., Kim-Rupnow, W. S., & Power, T. J. (2006). Video feedforward for reading. *The Journal of Special Education, 39*, 194–207. doi: 10.1177/00224669060390040101

Harrison, P. L., & Oakland, T. (2003). *Adaptive behavior assessment system for children* (2nd ed.). San Antonio, TX: Harcourt Assessment.

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy-research. *Journal of Consulting and Clinical Psychology, 59*, 12–19. doi: 10.1037/0022-006X.59.1.12

Lantz, J. F. (2005). *Using video self-modeling to increase the prosocial behavior of children with autism and their siblings*. Unpublished doctoral dissertation, Indiana University, Bloomington.

Lasater, M. W., & Brady, M. P. (1995). Effects of video self-modeling and feedback on task fluency: A home-based intervention. *Education and Treatment of Children, 8*, 389–407.

Miklich, D. R., Chida, T. L., & Danker-Brown, P. (1977). Behavior modification by self-modeling without subject awareness. *Journal of Behavior and Experimental Psychiatry, 8*, 125–130. doi: 10.1016/0005-7916(77)90032-5

Miklich, D. R., & Creer, T. L. (1974). *Self modeling as a behavior modification technique*. Oxford, England: Charles C. Thomas.

Prater, M., Carter, N., Hitchcock, C., & Dowrick, P. (2012). Video self-modeling to improve academic performance: A literature review. *Psychology in the Schools, 49*, 71-81. doi: 10.1002/pits.20617

Reynolds, C. R., & Kamphaus, R. W. (2004). *Behavior assessment system for children* (2nd ed.). San Antonio, TX: Pearson.

Schmitt, A. L., Livingston, R. B., Dykes, F., Scott, B. M., Galusha, J. M., Hua, S. B., & Garland, A. D. (2013). A theoretical model for measuring change: Reliable change index. *Research and Practice in the Schools, 1*(1), 8-14.

Article

Addressing Disruptive Behavior with Response to Intervention Practices: A Case-Study Approach to Improving Behavior in Student ADHD Populations

Jennifer N. Longoria
University of Texas at San Antonio

Billie Jo Rodriguez
University of Texas at San Antonio

Some state education agencies, districts, and schools have adopted evidence-based prevention and intervention models to address students' academic and behavioral challenges. Teachers are often responsible for gathering data on the prevalence of problematic behaviors, selecting and implementing interventions, and collecting follow-up data despite lack of formal training in these areas. The prevalence of Attention-Deficit/Hyperactivity Disorder (ADHD) has shown small but steady increases in school-age children since 1998, resulting in more students exhibiting academic and behavioral difficulties. This case study examines how school psychologists can use their behavioral expertise to assist teachers in selecting and implementing evidence-based interventions for students with externalizing behavior disorders, as well as interpreting the data for intervention effectiveness. Results of this AB research design found a contingency management intervention paired with positive behavior supports was effective in reducing disruptive behaviors. The case study also outlines how school psychologists can work within a consultation framework to support teachers with applied problem solving skills required in the RTI process.

Keywords: ADHD, consultation, RTI, school psychologist, single-subject design

The 2004 revision of the Individuals with Disabilities Education Act (IDEA) introduced the response to intervention (RTI) model as a strategy to prevent and address students' academic and behavior challenges. RTI models are based on a behavioral approach to improving performance by identifying relationships between target behaviors and interventions (Cooper, Heron, & Heward, 2007). Over the past decade state education agencies, districts, and schools have adopted evidence-based prevention and intervention models, such as RTI for school-wide positive behavior support, that incorporate instructional techniques in general and special education, progress monitoring, and decision-making in special education referrals. Evidence-based practices are

now considered the gold standard for addressing academic and behavioral challenges. Improved outcomes are dependent upon educators selecting effective interventions that pertain to the area of concern and implementing interventions with fidelity (Fixsen, Blase, Metz, & Van Dyke, 2013). Evidence-based practices are often limited by a lack of data collection, particularly while the intervention is in progress. This prohibits educators from assessing intervention effectiveness while it is being delivered and making necessary adjustments that could impro-

ve the outcome for the student (Becker & Domitrovich, 2011). Emphasis on using empirically supported strategies and data-based decision making in the school setting has resulted in an increased need for school psychologists to provide consultation services to educators.

Obstacles to Implementing RTI

Many Texas school districts have launched RTI models as a way to gather data on the prevalence and degree of academic and behavioral difficulties and to provide interventions matched to student need. As part of this model, teachers may be tasked with gathering baseline data on the academic or behavioral issues of concern and are sometimes even expected to independently select appropriate evidence-based interventions. For many educators, the process of documenting the problem, selecting and implementing interventions, and collecting follow-up data is daunting and beyond the scope of formal training (Castro-Villarreal, Rodriguez, & Moore, 2014). In many districts, educators are required to provide pre- and post-intervention data to assessment personnel, but they may lack formal training in such approaches (Kratochwill, Volpiansky, Clements, & Ball, 2007). Professional development trainings on RTI at the district level have emerged as a substitute for formal training on evidence-based practices, and although these experiences provide educators with an overview of RTI and data collection techniques, there is no substitute for systematic training at the university level (Kratochwill, 2008). School psychologists often support teachers as they become acquainted with RTI and help guide them through the processes outlined above. Despite local, state, and national promotion of RTI in legal and professional settings, implementation of RTI practices still faces challenges related to efficiency, validity, reliability, practicality, knowledge of strategies, and time constraints (Brown-Chidsey & Steege, 2010; Cook & Odom, 2013). Educators need continuous consultation support for evidence-based interventions to be implemented and assessed effectively (Forman, Olin, Hoagwood, Crowe, & Saka, 2009), and many school psychologists have the knowledge base to provide these services.

The Rise of ADHD in Student Populations

Response to Intervention has become integrated in educational systems as formal Attention-Deficit/Hyperactivity Disorder (ADHD) diagnoses in children have risen. ADHD is a neurodevelopmental disorder characterized by difficulty sustaining focus, shifting attention, and regulating or inhibiting behaviors across settings (American Psychiatric Association, 2013). Between 1998 and 2009, ADHD diagnoses in the United States increased 2.9 percent for boys and 1.9 percent for girls aged 5-17 (Akinbami, Liu, Pastor, & Reuben, 2011). The most recent figures on ADHD indicate 11% of children are diagnosed with ADHD, with boys being twice as likely to receive a diagnosis (Centers for Disease Control and Prevention, 2013). Compared to same-age peers without ADHD, students with the disorder demonstrate significant academic and social impairments across various domains (McConaughy, Volpe, Antshel, Gordon, & Eiraldi, 2013). Increasing rates of ADHD in school-aged children have resulted in increased consultation between school psychologists and teachers. Gathering information to support an ADHD diagnosis can be complex for school psychologists due to a variety of factors including limited background information, inconsistent reports from caregivers and teachers, and comorbidity with other disorders. Multi-tiered positive behavior support strategies have been found to be effective for students with externalizing disorders, such as ADHD, and providing this population with empirically supported strategies is critical for successful inclusion (Akin-Little, Little, Bray, & Kehle, 2009).

Benefits of Single-Subject Research Design in RTI

Single-subject research design has become increasingly relevant in educational research as educators aim to minimize academic and behavioral challenges that inhibit or disrupt learning processes. One of the most apparent benefits to using single-subject design is the opportunity to identify a functional relation between the behavior (dependent variable) and the intervention (independent variable) when the intervention is manipulated and to make predictions about when the behavior will occur (Casey et al., 2012). Predictions can be generated after reviewing

baseline data and verified by collecting data immediately after an intervention is removed. Removing an intervention and returning to baseline conditions will confirm if the changes in behavior during the intervention period can be attributed to the intervention (Kratochwill et al., 2010). Resuming the intervention after the second baseline period allows for further clarification on the relationship between the behavior and the intervention.

This case study demonstrates how school psychologists, designated LSSPs (Licensed Specialists in School Psychology) in Texas, can integrate research to practice techniques and provide support for teachers in the areas of data collection, intervention selection, and intervention implementation within the consultation framework. It also outlines a potential strategy for merging positive behavior support with evidence-based practices. The use of single-subject design models how school psychologists and teachers can assess the relation between the introduction of a behavioral intervention, medication, and a change in behavior for an individual diagnosed with ADHD. Training teachers on how to collect and interpret intervention data builds confidence for RTI implementation when future student challenges arise. Additionally, this case highlights how educators can use research design and data collection as part of the RTI process to tease apart the effects of behavioral interventions and medication, which results in valuable information on the relation between variables.

Purpose

The primary purpose of consultation was to determine if an evidence-based intervention incorporating positive behavior support strategies would be effective in reducing disruptive behavior for a first grade student. Emphasizing data collection and utilizing a single-subject research design was critical for determining if changes in behavior were linked to intervention alone or a combination of intervention and medication. A secondary goal was to provide training for the teacher in selecting and implementing RTI procedures including gathering supporting data to make decisions.

Method

Participants and Setting

Participants included a first year, first grade female teacher and a six-year-old Caucasian male (pseudonym Jared) from an elementary school located in a large suburb in Texas that was classified as a Title I-Bilingual campus. The student population (N=792) was primarily Hispanic (63%) but also included African American (19%), Asian (8%), Caucasian (8%), and Bi-Racial (2%) students. The first grade teacher was of Asian descent. The first grade classroom had 23 students, 13 female and 10 male. Five of the 23 students qualified for special education services. The teacher had student teaching experience in a behavior instruction classroom, a setting for students whose behaviors negatively affected their learning and required a reduced student-to-teacher ratio.

Jared was selected for participation during an instructional support team (IST) meeting in the first month of the school year to address behavior concerns. A review of informal data from the previous school year indicated Jared had a history of engaging in disruptive behaviors. Jared was receiving special education services for speech impairment in articulation. During the time of the school referral to the instructional support team, a private psychiatrist was also gathering information to determine if Jared met diagnostic criteria for ADHD and Oppositional Defiant Disorder (ODD). Jared was prescribed a 10mg dose of ADHD medication by the psychiatrist near the time the school-based interventions were implemented as a trial to see whether medication reduced inappropriate externalizing behaviors.

The LSSP Intern, a third year school psychology graduate student, consulted with Jared's teacher and assisted in the RTI process. The data were collected and analyzed as a university program requirement for completing a case study during the internship year.

Procedures

Baseline Data Collection

The LSSP Intern conducted two observations in the classroom setting as part of a functional behavior assessment. During the observations, Jared appeared

to ignore teacher directives and made inappropriate sounds and non-verbal gestures. He used inappropriate language, raised his voice above a typical talking volume, and threw classroom materials. Jared engaged in the problematic behaviors after receiving teacher directives that did not offer autonomy in decision-making, such as being told to transition to a new subject and when Jared wanted to avoid a task. It was hypothesized that the function of the behaviors was to escape the teacher's directive and gain her attention, possibly in an effort for her to acknowledge the various methods of defiance.

In an interview, the teacher described Jared's range of inappropriate behaviors and selected the three she believed to be most problematic and prevalent during instruction (i.e., non-compliance, inappropriate language, and destruction of materials). Inappropriate language was defined as using curse words, name calling, offensive slurs, sarcasm, and disrespectful statements (e.g., "I hate you."). Non-compliance was defined as instances when Jared did not follow instructions after two verbal commands. Verbal refusals (e.g., "No, I won't do that.") or non-verbal gestures (e.g., ignoring directions, pretending not to hear) were also recorded as non-compliance. Destruction of classroom materials was defined as damage to personal or classroom supplies including tearing, throwing, breaking, defacing, or any other action that impaired the function or value of an object beyond normal wear and tear. The LSSP Intern showed the teacher how to gather baseline data using a frequency data collection form. The teacher used the form to tally inappropriate language, non-compliance, and destruction frequency during each academic subject for five school days.

Baseline data collected by the teacher indicated Jared used inappropriate language a mean of 12.2 times per day (range 3 to 23). The teacher's expectation was that first grade students not use any inappropriate language in the classroom. Jared's non-compliant behavior was his most frequent, with an average of 35.4 instances each day (range 19 to 53). The expectation for first grade students was full compliance with teacher directives after no more than two verbal requests. The least common behavior during baseline was destruction, which averaged 1.4 instances per day (range 1 to 3). Destruction of school property was prohibited per the student code of conduct.

Intervention Description

Two interventions were put in place to help Jared gain positive attention from the teacher: The Morning Pep Rally and a Daily Behavior Report Card. Although the primary function of Jared's behavior was to avoid tasks, he also gained negative attention from the teacher when he refused to follow instructions. The teacher and LSSP Intern worked together to develop an intervention to increase positive interactions, and emphasis was placed on rebuilding the relationship between the teacher and student. The goal of the interventions was for Jared to receive non-contingent positive attention and develop a more collaborative relationship with the teacher to increase his compliance.

The purpose of the Morning Pep Rally intervention was to stage a positive interaction between Jared and teacher. Similar to an athletic pep rally, this intervention was an opportunity for the teacher to provide Jared with encouragement and non-contingent praise. The Morning Pep Rally took place before academic instruction began each day as students entered the classroom and awaited morning announcements. The duration was between one and two minutes. The teacher was encouraged to compliment Jared on at least one positive behavior from the previous day and discuss topics of interest (e.g., video games, after school activities) for Jared.

The Daily Behavior Report Card intervention utilized a token economy system that rewarded Jared for using appropriate language each class period. Interventions that use a classroom based token economy are empirically based, particularly with students who exhibit externalizing behavior difficulties (Kazdin & Weisz, 2003). The teacher was trained to assign Jared a rating based on a four-point likert scale at the end of each class period and verbally explain the rating to Jared, emphasizing positive performance. All ratings were recorded on the behavior report card by the teacher and totaled at the end of each day. If Jared met his pre-determined point goal, he was rewarded with a prize from the treasure chest. Jared and teacher signed his point sheet each day and sent it home for his parents to sign and return. Having Jared sign the point form served to foster personal responsibility for school performance. The parental signature served as an efficient way for the parent and teacher to communicate on the details of Jared's

daily performance between conferences. If Jared returned the form unsigned, the teacher was instructed to contact the parent and review Jared's performance via the telephone.

Jared also began pharmacological treatment with a well-known ADHD medication around the same time the interventions were implemented. There was some confusion in the school surrounding the exact day the medication trial began, but the parent confirmed it was initiated after the baseline data collection and the week the interventions began.

Intervention Implementation

After training the teacher on how to implement the interventions, the LSSP Intern met with Jared to complete a two-page reinforcement survey. The survey consisted of open-ended questions such as, "If I could do one thing with my teacher it would be…" and "The thing I like to do the most in my free time is…" Jared indicated interest in several rewards, most notably spending time on the computer and playing a game with his teacher. The consultant shared results of the reinforcement survey with the teacher, who added several of Jared's reward requests to his individual treasure chest. Jared was able to choose one reward when he met his daily point goal.

The teacher, Assistant Principal, and the LSSP Intern introduced the intervention to Jared as a game he would be playing to earn rewards he had indicated interest in on the reinforcement survey. Jared's parent was not available for the intervention introduction. The LSSP Intern contacted the parent following the intervention introduction to review each individual's role in the process and explained the importance of providing consistent responses to undesirable behaviors across settings. The parent was encouraged to present Jared with age-appropriate consequences (e.g., writing a brief apology letter to the teacher) in the home setting on days when Jared engaged in problematic behaviors outlined on the behavior report card.

Following five days of baseline data collection, both interventions were implemented for 10 school days. The initial goal required Jared to use appropriate language 33 percent of the day to receive a reward, which equated to earning a minimum of 22 points on the Daily Behavior Report Card. The target

percentage was derived after comparing daily trends of inappropriate language and set just above the current level of performance. Daily performance was calculated by dividing Jared's total accumulated points by the total points possible.

After the 11th day of intervention implementation, the interventions were discontinued, and the teacher measured the frequency of behaviors of concern using the baseline data collection form. This change in intervention measured whether the improvements in behavior were due to the trial dose of ADHD medication alone or to the school-based intervention. The intervention reversal was scheduled to occur for a total of three days, but Jared was absent once during the reversal period. During the two-day period of no intervention, problematic behaviors from the baseline stage reemerged even though Jared continued to take the ADHD medication. After removing the intervention for two days, the teacher returned to using the Daily Behavior Report Card and Morning Pep Rally.

Fidelity

Treatment integrity was assessed through direct observations, review of Jared's daily behavior report cards, and the intervention progress monitoring form. The teacher completed a daily progress monitoring form for each day the intervention was implemented. She recorded the meeting time and duration of the morning pep rally, which averaged 1.5 minutes. The teacher also recorded the total points Jared earned each day on the daily behavior report card and estimated Jared's non-compliant behaviors. The Intervention Monitoring forms were completed for 90% of the intervention period.

Results

Comparison of pre- and post-intervention data revealed a reduction in non-compliance, inappropriate language, and destruction of classroom materials. Figure 1 displays that after five days of intervention, instances of non-compliance showed the most reduction in frequency ($M = 7$) when compared to the first baseline period ($M = 35.4$). The frequency of inappropriate language also decreased from baseline ($M = 12.2$) to the second intervention ($M = 3$) period. Destruction of materials was the most infrequent of
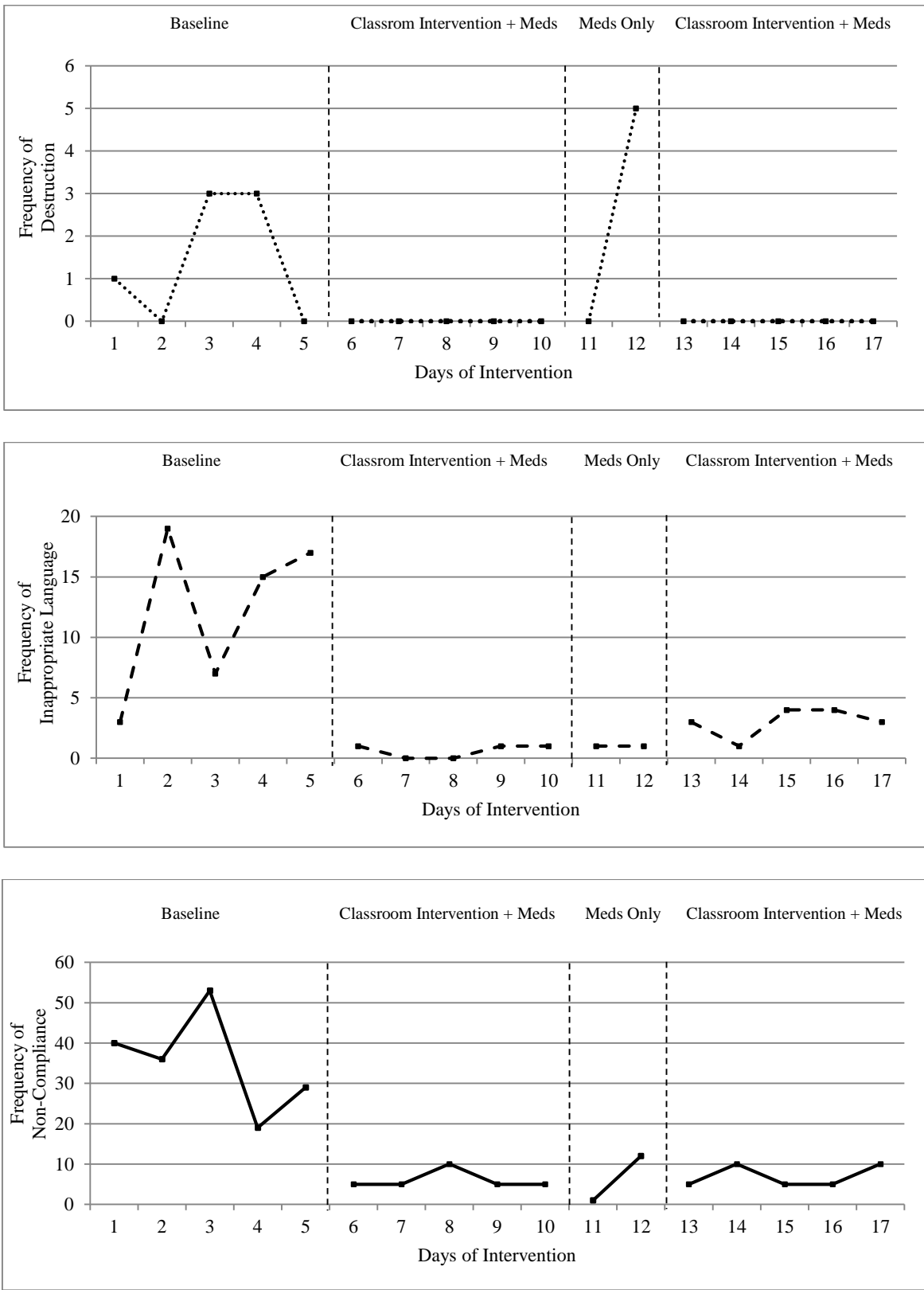
Figure 1. Problematic Behaviors Frequency Pre- and Post-Interventions

the three target behaviors at baseline (M = 1.4), and Jared exhibited no destructive behaviors (M = 0) after the intervention reversal.

Data from the two days the teacher stopped implementing the intervention revealed Jared began resuming the behaviors of non-compliance (M = 6.5), inappropriate language (M = 1), and destruction (M = 2.5). It appeared that Jared maintained appropriate behavior during the first day of reversal, possibly due to conditioning, but he quickly resumed the disruptive behavior when the reward was withheld. During the reversal period, problematic behaviors increased in frequency to levels not observed during the classroom intervention, despite the fact Jared continued pharmacological treatment. The rebound in disruptive behaviors following the first intervention trial suggests the interventions, not the medication alone, were responsible for the decrease in inappropriate language, non-compliance, and destruction of materials.

Discussion

Results from this case study provide additional evidence that positive behavior approaches, combined with a token economy system, are effective strategies for teaching new skills to students and reinforcing socially appropriate behavior. Data from the intervention reversal period indicated the problematic behaviors displayed at baseline were not mediated by medication alone and resumed when the intervention was discontinued. The return of Jared's problematic behaviors while taking medication confirmed to the teacher and LSSP Intern that the combination of both interventions was successful in reducing Jared's three most problematic behaviors and improving relations with the teacher.

The Daily Behavior Report Card intervention was selected due to empirical support linking token economies with behavioral improvements in student populations, particularly those with externalizing behaviors (Kazdin & Weisz, 2003). The Morning Pep Rally intervention was incorporated to reinforce unconditional positive interactions between Jared and the teacher, as outlined in positive behavior support strategies and aligned with one function of Jared's problem behavior (i.e., obtain teacher attention). Data from this case study support previous research

that has found teacher-mediated interventions are effective in reducing externalizing behaviors (Akin-Little et al., 2009).

This case study outlines how school psychologists can apply their formal research training to evaluate the effectiveness of evidence-based intervention strategies, particularly with students who receive pharmacological treatment. Some students treated for ADHD with medication may require additional intervention support, and collecting data is paramount in identifying what strategies are effective in supporting prosocial behavior. Practitioners can also use the consultation process as an opportunity to support educators in implementation of foundational RTI practices, such as data collection and behavior analysis, as well as in selection, implementation, and evaluation of interventions. Lack of formal training and brief exposure to RTI practices in professional development activities leave many educators unprepared to initiate and execute RTI procedures when addressing behavioral concerns (Kratochwill et al., 2007; Kratochwill, 2008). Without appropriate training, educators cannot be expected to collect data that are reliable and valid for decision-making purposes, such as special education referrals (Brown-Chidsey & Steege, 2010). Consultation with school psychologists can serve as an informal and applied training opportunity for new or veteran educators who lack experience with emerging RTI practices or require assistance in specific areas of practice. The consultation process can also strengthen working relationships between school psychologists and teachers and prepare educators to implement RTI procedures independently when future academic or behavioral concerns arise.

Limitations

The case study illustrates how single-case design research can be used in practice to inform intervention implementation, but the results of this case study are limited due to several factors. First, the single student and teacher participant dyad in this study makes generalization to other populations difficult. Although we employed a reversal design in an attempt to strengthen the results, this reversal was limited by two factors. The most significant limitation is that different types of data were collected during the re-

versal period from the initial baseline, making it difficult to interpret seamlessly. There was also a lack of reversal data compared to baseline and intervention frequency data. This results in an AB case study design, which shows a correlation but cannot be used to draw causal effects. Another limitation is that the consultant who was not blind to the study conducted observations, and inter-observer agreement data were not gathered.

Other considerations include the interventions required increased time commitment from the teacher in the mornings when she was tasked with other responsibilities and throughout the day as she assigned a performance rating at the end of each academic period. Results of this case study may have been different if the teacher was tasked with providing intervention support to multiple students in the classroom or if the teacher's overall classroom behavior management strategies were less effective. The intervention incorporated increased time and support from the child's parent in the form of providing praise for use of appropriate language, compliance with teacher directives, and respecting classroom materials, or presenting age-appropriate consequences when the student failed to demonstrate the abovementioned behaviors in the school setting. Although the parent verified receiving information about daily performance by signing the behavior report card, it remains unknown if the parent modeled prosocial behavior in the home setting or consistently delivered age-appropriate consequences.

Implications & Future Research

This case study exhibits how school psychologists can incorporate research and single-case design methodology into RTI practices to enhance data-based decision making. It highlights the potential positive outcomes for both students and teachers that can be obtained when applying single case research design to an educational setting. Additionally, the study illustrates the feasibility of this practice, as a LSSP Intern using only the resources available in a typical educational context conducted all methods employed in the current case study. Future research should emphasize selecting more rigorous designs that will allow for more consistent measures of data collection across baseline and intervention periods. Determining the method and measurement criteria

for target behaviors beforehand will reduce the complexity of data analysis and interpretation. Practitioners should capitalize on academic and behavioral consultation cases by training educators in RTI practices, which will strengthen the school's RTI system over time and ultimately improve outcomes for students in those systems.

References

Akinbami, L. J., Liu, X., Pastor, P. N., & Reuben, C. A. (2011). *Attention deficit hyperactivity disorder among children aged 5-17 years in the United States, 1998-2009.* NCHS data brief, no 70. Hyattsville, MD: National Center for Health Statistics.

Akin-Little, A., Little, S. G., Bray, M. A., & Kehle, T. J. (2009). *Behavioral interventions in schools: Evidence-based positive strategies.* Washington, DC: American Psychological Association.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.

Becker, K. D., & Domitrovich, C. E. (2011). The conceptualization, integration, and support of evidence-based interventions in the schools. *School Psychology Review, 40,* 582-589.

Brown-Chidsey, R., & Steege, M. W. (2010). *Response to intervention: Principles and strategies for effective practice* (2nd ed.). New York, NY: Guilford.

Casey, L. B., Meindl, J. M., Frame, K., Elswick, S., Hayes, J., & Wyatt, J. (2012). Current trends in education: How single-subject research can help middle and high school educators keep up with the zeitgeist. *The Clearing House, 85,* 109-116. doi:10.1080/00098655.2012.655344

Castro-Villarreal, F., Rodriguez, B. J., & Moore, S. (2014). Teachers' perceptions and attitudes about response to intervention (RTI) in predominately Hispanic serving schools: A qualitative analysis. *Teaching and Teacher Education: An International Journal of Research and Studies, 40,* 104-112.

Centers for Disease Control and Prevention. (2013). Attention-deficit/hyperactivity disorder (adhd). Retrieved from http://www.cdc.gov/ncbddd/adhd/data.html

Cook, B. G., & Odom, S. L. (2013). Evidence-based practices and implementation science in special education. *Exceptional Children, 79*(2), 135-144. doi:10.1177/001440291307900201

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2nd ed.). Upper Saddle River, NJ: Pearson Education.

Fixsen, D., Blase, K., Metz, A., & Van Dyke, M. (2013). Statewide implementation of evidence-based programs. *Exceptional Children, 79*(2), 213-230. doi:10.1177/001440291307900206

Forman, S. G., Olin, S. S., Hoagwood, K. E., Crowe, M., & Saka, N. (2009). Evidence-based interventions in schools: Developers' views of implementation barriers and facilitators. *School Mental Health, 1,* 26-36.

doi:10.1007/s12310-008-9002-5

Kazdin, A. E., & Weisz, J. R. (2003). *Evidence-based psycho-therapies for children and adolescents.* New York: Guilford.

Kratochwill, T. R. (2008). The good, the bad, and the ugly: A report card on evidence-based practices in the schools. *Communiqué, 36,* 1-14.

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

Kratochwill, T. R., Volpiansky, P., Clements, M., & Ball, C. (2007). Professional development in implementing and sustaining multitier prevention models: Implications for response to intervention. *School Psychology Review, 36,* 618-631.

McConaughy, S. B., Volpe, R. J., Antshel, K. M., Gordon, M., & Eiraldi, R. B. (2011). Academic and social impairments of elementary school children with attention deficit hyperactivity disorder. *School Psychology Review, 40,* 200-225.

Article

The Effects of a Parent Delivered Video Modeling Intervention to Improve Generalized Transition Skills in a Child with Autism

Jacklyn E. Pierce and Ginger L. Kelso
Stephen F. Austin State University

Many children naturally develop transition skills, but children with autism often struggle to cope with change in their environment caused by transitions between activities. This can result in problematic behaviors and wasted instructional time at school. However, video modeling is a popular technique that may result in effective and efficient interventions for transition skills. In video modeling, the student watches a video of the target behavior being completed appropriately prior to engaging in that behavior. In this study, an 11-year-old girl with autism used video modeling to improve transition behavior in the home while generalized effects of this intervention were measured in the school. Results show that video modeling alone was insufficient to produce improved transition skills in either setting. However, the addition of an embedded rule in the video model resulted in skill acquisition. This study demonstrates how a brief intervention can produce generalized behavior change.

Keywords: transitions, transition skills, intervention, video modeling

Throughout the day, children engage in a variety of activities. Most children develop the skills needed to move from one activity to another with ease. These movements between activities are referred to as transitions. School-age children are expected to walk between classes appropriately, turn off the computer when asked, and put away materials when an activity has ended. Students also must participate in transitions that may be unexpected or outside of the daily routine such as campus fire drills or assemblies. In the home, children must transition among activities such as meal times, homework, TV time, and bedtime. However, children with autism spectrum disorders often have difficulty transitioning from one activity to another and may perseverate on engagement in a favored activity (Heflin & Alaimo, 2007). This necessitates the direct teaching of transition skills in this popula-

tion. The American Psychiatric Association states that individuals with autism often struggle in situations that involve change in the environment (American Psychological Association [APA], 2013), which is inherent in the need to move from one activity to another. Moreover, the teaching of such transition skills is complicated by the tendency for children with autism to develop behavior problems when faced with the need to transition to a new activity.

According to the Diagnostic and Statistical Manual of Mental Disorders (DSM-5), individuals with autism are characterized by "insistence on sameness, inflexible adherence to routines, or ritualized patterns of verbal or nonverbal behavior" including "extreme distress at small changes" and "difficulties with transitions" (APA, 2013, p. 50). Difficulties in appropriately responding to environ-

mental change and navigating transitions can disrupt many daily routines at home and at school. Individuals with autism are also characterized by a "strong attachment to or preoccupation with unusual objects" and "excessively circumscribed or perseverative interests" (APA, 2013, p. 50). This can further complicate transitions when a child is asked to stop an activity with a favored object and begin a new activity with different materials.

Sterling-Turner and Jordan (2007) reviewed literature concerning various intervention techniques designed to improve transition skills for individuals with autism. Among the interventions described as successful for individuals with autism are verbal techniques such as verbal or auditory cues, behavioral momentum, and visual techniques such as pictorial cues and activity schedules. Sterling-Turner and Jordan also describe the use of video priming or modeling as a useful technique to improve transition skills. Video modeling is a technique that has been successfully used to improve a variety of skills for children with autism (Darden-Brunson, Green, & Goldstein, 2008). In a video modeling intervention, the child watches a video of the desired behavior being performed appropriately and independently. Darden-Brunson et al. describe various methods for creating video models in which an adult, a peer, or the target child him/herself serve as the model. Alternatively, the video may show a point-of-view perspective in which no model is visible. Instead, the environment is viewed from a first-person perspective as the target child would see it. Video modeling is intended to increase the accurate performance of the behavior being viewed on the video. This technique is thought to benefit children with autism because it incorporates visual learning strategies, which have been noted as an area of strength for this population (Cihak, Fahrenkrog, Ayers, & Smith, 2010).

Over the previous decade, several reviews of literature have been published documenting the effects of video modeling specifically for children with autism. Ayers and Langone (2005) showed that video modeling has successfully improved a variety of social and functional skills such as conversations, grocery shopping, and food preparation. Delano (2007) extended this evaluation to show that video modeling can improve skills in four behavioral categories: social-communicative behaviors, functional living skills, perspective taking skills, and challenging behaviors. Delano also concluded that skills learned through video modeling often generalized to untrained conditions. McCoy and Hermansen (2007) evaluated the effects of the types of models used in video modeling interventions for children with autism. They divided studies into categories based on whether the video model incorporated adults, peers, self, point-of-view, or mixed models. McCoy and Hermansen found that peer and self-models are more effective for children with autism when compared to adult models. Sherer et al. (2001) compared peer and self-modeling to teach conversational skills. Because the effects varied among the children, the authors concluded that peer and self-modeling were equally effective. Shukla-Mehta, Miller, and Callahan (2010) evaluated whether studies demonstrating effects of video modeling for children with autism met the criteria for evidence based practices as outlined by Odom et al. (2005). They found that video modeling for children with autism when used to teach social and communication skills has some evidence to support its effectiveness.

More recently, Tullis, Cannella-Malone, and Payne (2014) reviewed interventions for between-task transitions for individuals with developmental disabilities including autism. Of the 32 studies reviewed, several antecedent and consequence-based interventions were identified to address transition skills for individuals with developmental disabilities. However, only four studies included video modeling as an intervention for individuals with autism. One of these studies (Schreibman, Whalen, & Stahmer, 2000) focused on decreasing challenging behaviors during transitions, while the other three studies targeted the teaching of transition related behaviors (Cihak, 2011; Cihak et al., 2010; Mechling & Savidge, 2011). All four studies indicated positive results for video modeling, but only one of the four studies included generalization data. Since the publication of Tullis et al. (2014), one additional study has also been added to the literature base on video modeling for improving transitions (Spriggs, Knight, & Sherrow, 2014). Each of these five studies will be reviewed.

Schreibman, Whalen, and Stahmer (2000) used a form of video priming to reduce problem behaviors for three children with autism during transi-

tions. In this study, video priming consisted of showing the child a video of the transition environment, but no model was included in the videos. These videos would be characterized as "point-of-view" instead of adult, peer, or self-modeling. Schreibman et al. found that all three children decreased challenging behaviors during intervention. They also found that the improvements in behavior generalized to two additional conditions: familiar routine with no video and familiar routine with irrelevant video. Cihak et al. (2010) examined the use of video modeling to increase independent transition skills for four elementary students with autism in the general education classroom. The study used a single subject reversal design to monitor ten transitions daily for each student. Transitions were considered to be independent and successful only if the student transitioned without any form of prompt or assistance from the teacher and engaged in no problem behavior. In baseline children transitioned independently during 30% or less of opportunities. Upon introduction of video modeling, children watched a video on a video iPod that showed a positive self-model of the child transitioning independently from one location to another. During intervention all children increased independent transitions to the criterion of 100% accuracy for three consecutive sessions. All four students maintained independent transitions at or above 90% nine weeks after the intervention ceased.

In another study, Cihak (2011) compared the relative effects of two visual strategies to improve transition skills for children with autism. Four middle school-aged children with autism received both static-picture schedules and video based schedules. The static-picture schedules included pictures of the student transitioning between each task. The video models incorporated both self-models as well as point-of-view modeling showing transitions. In this study, both static-picture schedules and video based schedules resulted in improved transitions for all children. However, when the results of pictorial and video schedules were compared, the intervention that resulted in the quickest improvement varied between children.

Mechling and Savidge (2011) also combined video and picture schedules to improve transitions. In their study, three middle school students with autism completed four tasks followed by a rein-

forcement period. To improve task completion as well as transitions within and between tasks, the students were provided with a Personal Digital Assistant (PDA) that included number, picture, auditory, and video prompts. The students could choose which level of prompt was needed to successfully complete the tasks and transition to the reinforcement period. Within-task transitions required students to complete one activity and begin the next. The video models used during these transitions did not show transition behavior but instead focused on correct task completion. However, video models for between-task transitions showed the point of view perspective of a student moving and accessing the reinforcement provided. All students improved transitions between the task and reinforcement period. During between-task transitions, the students chose the type of prompt needed by selecting icons on the PDA. They could choose between a picture of the reinforcing item, an auditory instruction to get the item, or a video showing a first-person perspective of locating the item. All students chose the video prompt most often. Although video prompts were successful for increasing transitions to reinforcement, it is not clear whether the video was necessary. It is possible that the students could have completed the tasks and transitions successfully with picture or auditory prompts.

Spriggs et al. (2014) conducted a similar study in which visual activity schedules were combined with video models. However, in this study none of the video models showed transition behavior. The students used a visual activity schedule to identify the next task to be completed, and the video model showed the steps required to complete the task. Although high school students with autism improved transitions in this study, it is unclear whether the visual activity schedule alone would have produced similar effects for transitions.

These studies provide initial evidence that video modeling may be an effective intervention for improving transition skills for children with autism. The existing studies highlight a few specific areas for future researchers to explore. First, all five studies focused on difficulty with transitions as indicated by demonstration of inappropriate behavior or refusal to transition. However, in four of these studies no attempts were made to analyze the types of activities that were experienced just before and just

after the transition. Therefore, it is unknown if the activities that the students were leaving or transitioning to were preferred activities. An exception can be found in Mechling and Savidge (2011) as students transitioned from a curricular task to a reinforcement period. It is assumed that the reinforcement was a preferred activity. Whether a child is being asked to stop or start activities that they find enjoyable or aversive could moderate the effects of a video modeling intervention.

The nature of the activities included in the transition also brings into question the timing of the video model. Three of the four studies required the children to watch the video after the previous activity had been stopped. For example, in the study by Cihak et al. (2010), the children watched the video just prior to being given the instruction to transition. Cihak (2011) gave the instruction to transition, the child watched the video, and then the child moved to the next activity. Although watching a video model just before the modeled behavior is expected is typical of a video modeling intervention, it creates some difficulties when applied to transition behaviors. The child is actually transitioning from an activity, to watching a video, and then to another activity, which changes the nature of the transition requirement. Schreibman et al. (2000) required the children to watch the video before entering the transition environment. For example, the child watched a video before leaving home while the transition requirement was to move from one store to another within a mall. This study showed improvements in transition behaviors even though the video model was temporally removed from the transition behavior.

Additionally, it has been shown that video models often lead to generalized behavior changes for children with autism (Darden-Brunson, Green, & Goldstein, 2008). However, more research is needed to show whether this is also true when video models are used for transition skills. Schreibman et al. (2000) showed that improvements in transition skills generalized to no video and irrelevant video conditions. However, more research is needed to determine whether behavior changes obtained in one setting are generalized to other settings or to other transition behaviors.

Therefore, the purpose of the present study was to explore these aspects of video modeling as applied to transitions. This study evaluated the effects of video models when the child is required to transition from a preferred to a non-preferred activity, since this is a situation that is likely to produce problem behavior or refusal to transition. This study also examined whether a video model that is watched several minutes before a transition will produce the same positive effects as those in previous studies that were watched immediately before the transition. Finally, this study assessed the effects of using video modeling in a home setting to improve transition skills in that setting while also monitoring the generalization of transition skills to the school setting.

Method

Participant

One 11-year-old female diagnosed with autism participated in this study. To protect confidentiality, she was given the pseudonym Terrie. Terrie attended a public elementary school in a rural area daily and received in-home training provided by the school district as a part of the state required autism services. She was selected to participate in this study based on her need to improve transition behaviors as well as her participation in the in-home training. Terrie was able to communicate in short phrases and could follow one step directions. When asked to transition to a non-preferred activity, Terrie often made vocal sounds not resembling words such as grunts or screams. The Institutional Review Board approved this study and parental informed consent was obtained prior to participation in this study.

Terrie's mother and classroom teacher also participated in this study. Terrie's mother implemented the intervention at home and collected data in the home setting. Terrie's mother had no formal training in autism intervention. However, she regularly attended state level conferences on the topic of autism. Terrie's classroom teacher collected data in the school setting. The teacher had taught special education classes for 14 years and held teaching certificates in special education and general education.

Settings

Intervention took place at the child's home in the living room, kitchen, and bathroom. The kitchen was approximately 120 square feet and included a sink, oven, refrigerator, and three feet of counter space. The living room was located adjacent to the kitchen and was approximately 250 square feet with a front door entrance to the house, two couches, and a television. A small circular table with four chairs was located in the area between the kitchen and the living room. The bathroom was approximately 120 square feet with a bathtub, sink, and toilet. Generalization data were collected at the child's school. The child's classroom was approximately 300 square feet with one rectangular table, two individual desks, teacher desk, and four bookshelves. The classroom had an individual bathroom and exercise corner including a treadmill, balls, and bench. The class consisted of two teachers and approximately three other students.

Materials

Materials for the study included an iPad 2 with video capabilities, coloring book with markers, math and spelling work sheets, and the student's lunch kit. The iPad was provided by the student's school district. Video recording was conducted using the Apple Video application. Terrie was recorded coloring in her book and was told, "It is time to do your work and complete your math sheet." Terrie put her materials away and began working on her math worksheet. The recording continued as she worked on her worksheet for five seconds. The video used for this study was edited to show only appropriate transition behaviors. Although the specific activities engaged in before and after the transition differed throughout the day, the video showed only one example of transition behavior from coloring to math.

Data Collection

Data collection sessions were alternated between home and school settings. Data were collected at school on the next available school day following a home session. The primary data collector in the home was the mother of the participant. In the school setting, the teacher served as the primary data collector. Data were collected daily for each of five transitions to non-preferred activities. Data were collected using paper and pencil behavior monitoring forms provided to the mother and teacher. Each form included a space to indicate a plus or minus for each of the five transitions. A plus indicated an appropriate transition while an inappropriate transition was indicated by a minus sign. The first author verbally described the behavior definition and data collection techniques to the parent and teacher. No additional training was provided. All data collectors wore watches with second hands to track the passage of time for the purpose of data collection.

The first author also collected data along with the primary data collectors on 20% of baseline and intervention sessions in both the home and school settings to determine interobserver agreement. Interobserver agreement resulted in 100% agreement between observers for both home and school settings. An agreement was defined as both observers recording either an appropriate or inappropriate response for each of the five transitions.

Dependent Variable

The target behavior during this study was appropriate transitions. An appropriate transition was defined as the child ceasing a preferred activity and beginning a non-preferred activity without making any inappropriate non-word vocal sounds. Closed-mouth humming or saying recognizable words was considered appropriate for the purpose of this study. Additionally, to be considered an appropriate transition, the child must begin the non-preferred activity within 30 seconds of being given the verbal instruction to transition. Data were collected in both the home and school settings. However, the intervention was only implemented in the home setting. Data collected in the school were for the purpose of assessing generalization across settings.

Design

This study was conducted in the form of a single subject reversal design. Baseline data were collected for six sessions in order to establish a sta-

ble pattern of behavior. Intervention implementation was scheduled until the student reached the criterion of at least 80% appropriate transitions at school and home for three consecutive sessions. However, due to the failure to reach criterion, a second intervention phase was added to this design. After meeting the criterion within the second intervention, the intervention was withdrawn for four sessions. Finally, following the withdrawal of intervention, the second intervention was reinstated until the student again reached the criterion of 80% for three consecutive sessions.

Procedures

General Procedures. During each session at home and at school, Terrie was instructed by her teacher or mother and given five opportunities to transition from a preferred activity to a non-preferred activity. Various activities were used for these transition opportunities. Preferred and non-preferred activities were selected based on parent report. Preferred activities were those that Terrie generally engaged in without resistance. Non-preferred activities were those in which she generally exhibited problem behavior when instructed to engage. Preferred activities consisted of coloring and watching TV at home. However, only coloring was used for school transitions. Non-preferred activities in the home included activities such as math worksheets, using the restroom, washing hands, cleaning out a lunch kit, or writing spelling words. These same non-preferred activities were used at school with the exception of cleaning out the lunch kit.

During each session, the parent or teacher instructed Terrie to transition from a preferred activity to a non-preferred activity. If she did not begin the non-preferred activity within 15 seconds, the teacher or mother verbally prompted Terrie to begin the non-preferred activity. If an appropriate transition did not occur within 30 seconds, the transition was considered inappropriate, and the teacher or mother removed the preferred activity that Terrie was engaged in (coloring book or TV). Any tantrum behavior such as screaming or throwing materials was ignored by the teacher or mother. Upon completion of non-preferred activities, the mother or teacher provided verbal praise and allowed Terrie to

return to a preferred activity. Praise was available during both baseline and intervention conditions. Regular in-home training targeting goals unrelated to transitions continued throughout all phases of this study.

Baseline. During baseline, after Terrie engaged in a non-preferred activity for approximately 10 minutes, the mother or teacher indicated that it was time to transition by giving a verbal instruction to begin the non-preferred activity. For example, when the non-preferred activity was math worksheets, the mother or teacher said, "It is time to do your work and complete your math sheet." Exact wording of the prompt was adapted by the mother or teacher to sound natural based on the type of transition required. No video model was implemented during this phase of the study.

Video Modeling. Procedures during the video modeling intervention were identical to baseline with the addition of a video model at the beginning of each session in the home setting only. No intervention was implemented in the school. During the video modeling intervention, the mother began each session by presenting the video on the iPad to Terrie. After the video was complete the child was allowed to engage in a preferred activity for about 10 minutes before the mother told Terrie, "It is time to do your work and _____." Again, the exact wording was adapted to sound natural based on the type of transition required.

Success during video modeling was determined based on a criterion. The video modeling intervention was scheduled to continue until Terrie reached 80% appropriate transitions in both home and school settings for three consecutive sessions. However, after six sessions of video modeling, Terrie failed to meet the established criterion, therefore, the intervention was altered to increase Terrie's likelihood of success.

Video Modeling + Rule. The video modeling + rule intervention was identical to the video modeling intervention with the addition of a verbal rule embedded in the video model. Because the target behavior depicted in the video was somewhat subtle (remaining quiet while transitioning), it was decided that a verbal rule describing the target
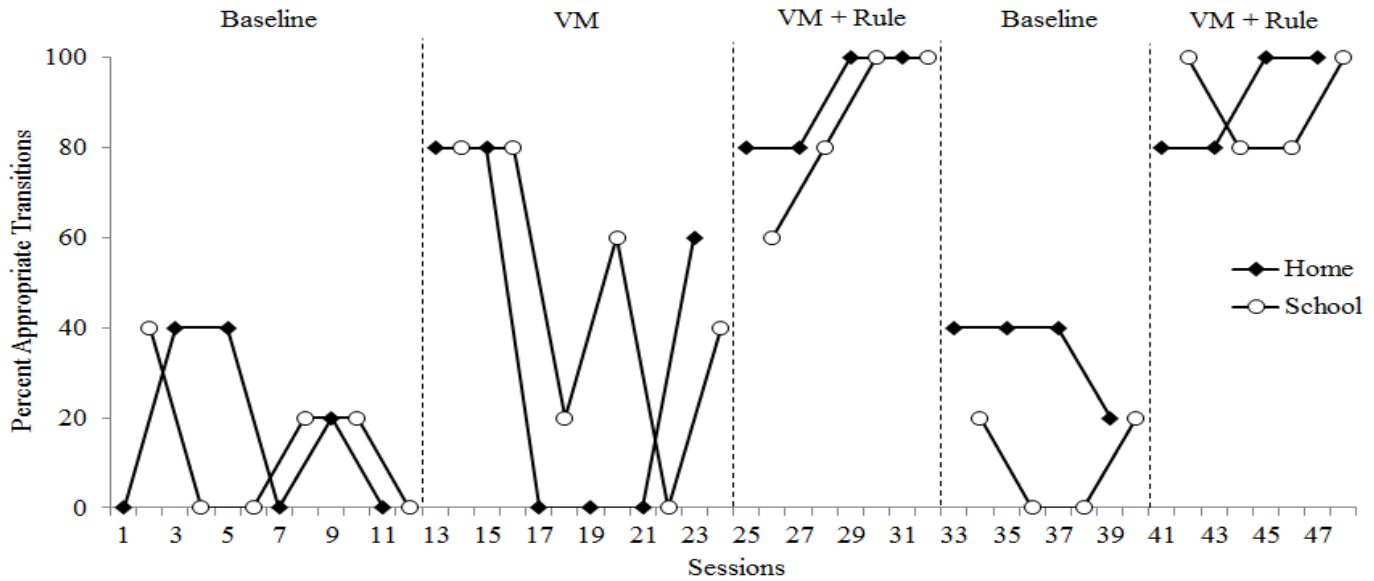
Figure 1. Percent of appropriate transitions at home and school.
VM: Video modeling

behavior might draw Terrie's attention to the relevant features of the video model. The rule stated, "When you are told to do your work, you do your work with a quiet mouth." This spoken rule was recorded and added to the video model just before the video of the student transitioning began to play. Again, this intervention was considered successful if Terrie achieved 80% appropriate transitions in both home and school settings for three consecutive sessions. Once criterion was met, the intervention was withdrawn for four sessions and then reinstated until the criterion was met for a second time.

Results

Results for this study are displayed in Figure 1. For each session, both home and school data points are displayed. The school session occurred on the next available school day after the home session. During baseline at home and at school, Terrie performed between 0-40% appropriate transitions for six sessions of baseline. Five out of the six data points at school were 20% and below. It should be noted that there was a two week break between sessions three and four due to a school holiday.

During the video modeling intervention, Terrie performed 80% appropriate transitions dur-

ing the first two consecutive sessions in the home setting. The following four sessions were inconsistent and ranged from 0-60% appropriate transitions with three consecutive sessions of 0% appropriate transitions. At school, Terrie performed 80% appropriate transitions during the first two consecutive sessions, which was the same rate as achieved at home. Also, similar to the home settings, the following four days at school were inconsistent and ranged from 0-60% appropriate transitions. Due to failure of reaching mastery criterion for video modeling, the intervention was modified to include the verbally stated embedded rule.

During the video model + rule intervention, Terrie performed 80 – 100% appropriate transitions at home for four consecutive sessions. At school, Terrie gradually increased performance from 60-100% appropriate transitions within four sessions with the last three sessions remaining at 80% or better. Criterion was reached for the video modeling + rule intervention.

Following the achievement of criterion during the video modeling + rule intervention, the intervention was withdrawn to allow the behavior to return to levels similar to baseline. During this second baseline phase, Terrie performed 20-40% appropriate transitions for four sessions at home and 0 – 20% appropriate transitions at school.

Following the second baseline phase, the video modeling + rule intervention was reinstated. This intervention continued until the criterion was again met. Terrie performed at 80-100% appropriate transitions for four sessions. Criterion was again achieved, and the study procedures were concluded.

Discussion

The results of this study provide a partial replication of the effects of video modeling as an effective intervention for transition skills for children with autism. Unlike previous studies, the child in this study needed additional support to successfully transition from a preferred activity to a non-preferred activity. This support was provided in the form of a verbal rule embedded in the video model. However, the lack of effect for the video model alone may have been due to the abstract nature of the behavior required. The student may not have attended to the "quiet mouth" aspect of the video model. The verbal rule may help the child attend to the critical features of transitions being demonstrated in the video model.

The effects of embedding a rule in a video model were demonstrated by Apple, Billingsley, and Schwartz (2005). The researchers created video models of children responding to others by giving compliments. Following the video of peer models, an adult stated the rule that described the desired behavior. This study showed that video modeling with an embedded rule successfully increased compliment-giving in response to another person's initiation of an interaction. However, the authors state that no comparison was made between video models with and without rules. Therefore, no conclusions can be drawn about which component of the video model resulted in improved performance.

The current study adds to the literature on video modeling with embedded rules in several ways. In the current study, the embedded rule was included at the beginning of the video instead of at the end. The effects of a video model with embedded rule were also extended to transition skills, which are topographically and functionally different than compliment-giving. Also, the current study demonstrates that video modeling with an embedded rule was effective after the same video without

the rule had shown positive but inconsistent results. Future researchers should further evaluate the relative effects of video modeling alone and video modeling with embedded rules.

The positive, yet inconsistent, results of using video modeling alone may point to the added difficulty of requiring the student to transition from a preferred to a non-preferred activity. Previous uses of video modeling for transitions did not specify the preference of the activities that preceded and followed the instruction to transition. This is an area of research that needs further development to determine if video models can produce consistently positive effects when the child must leave a preferred activity and begin a non-preferred activity.

Another potential explanation for the less than consistent results of video modeling alone may involve the timing of the video model. Video models are typically watched immediately prior to performing the desired behavior. However, to preserve the nature of the transition, the video model was watched approximately 10 minutes prior to the transition. This added delay could have resulted in inconsistent responding. Previous researchers (Schreibman et al., 2000) have shown positive effects of video priming when the video was watched at a time prior to the transition requirement. However, the length of time that passed between watching the video model and performing the transition was not reported. In the current study, it is unclear whether the time between video model and behavioral expectation was too long. The timing of the video model needs to be further explored. Although it has been shown that inserting the video model within the transition is effective in producing successful transitions (Cihak et al., 2010; Cihak, 2011), it disrupts the natural transition sequence. The current study shows that, with the addition of an embedded rule, video modeling successfully improved performance after a 10 minute delay. Improved performance at school during a subsequent session also shows that video models may improve performance after an extended time delay. However, more research is needed to determine whether it is possible to obtain consistently positive effects of video models when the video is shown at a time prior to the transition.

This study also replicates the results of previous video modeling research in that the effects of

intervention generalized to an untrained setting. In this study, the video model + rule intervention only occurred in the home, but mastery of the criterion occurred in both home and school settings. This demonstrates the potential efficiency of video modeling as an intervention to improve transition skills in multiple settings. However, in the current study, the transitions that occurred at home and at school involved similar activities and materials. It is possible that generalization would not be as successful if the transitions were of a more dissimilar nature. Further research is needed to determine the parameters of generalization between transitions involving very different tasks or materials.

Other factors may also have moderated the effects of the intervention in this study. During both video modeling and video modeling + rule interventions, it was noted that Terrie was consistently unsuccessful during transitions that required non-preferred activities that involved using the restroom or washing her hands. However, during sessions 15 and 16, Terrie appropriately transitioned to the restroom. Later, the mother reported that during these two sessions she instructed the student to use the restroom and added "then you can watch TV." This resulted in 100% of appropriate transitions on those sessions as compared to 80% appropriate transitions during the two previous sessions. The added reinforcer of watching TV deviates from the original intervention and is a factor that could explain the improved performance in transitioning during those two sessions. However, the student was performing at criterion levels prior to the offer of an additional reinforcer. Additionally no modifications were made to the procedures at the school on those days, yet Terrie still met criterion at school. Terrie also met criterion during the second implementation of this intervention following a brief withdrawal of intervention procedures. No deviations from intervention procedures were noted during the second implementation phase. Therefore, the improved performance of transitions can be attributed to the video modeling + rule intervention. However, future studies should include training and monitoring to ensure proper implementation of procedures by all interventionists.

The current study also adds to the existing literature on video modeling by incorporating video modeling into an in-home parent-implemented in-

tervention. Children with autism often receive in-home training, which involves parents as the interventionists. Video models are particularly appropriate for use by parents because no specialized training is required to use this intervention and implementation does not require excessive time or effort on the part of the parent (Darden-Brunson et al., 2008). This study provides evidence that the delivery of an intervention in the home setting by a parent can improve skills at home and at school.

In this study, the types of non-preferred activities required during transitions varied from day to day based on the types of skills required in the home or school that day. This means that a variety of academic as well as independent living skills were included throughout this study. The variability of non-preferred activities may have led to some of the variability in data from session to session. More research is needed to determine whether the types of non-preferred activities may modify the effects of video modeling interventions. In the early stages of this study, the child consistently failed to transition appropriately to non-preferred activities that involved the bathroom such as using the toilet or washing hands. However, these patterns may be idiosyncratic.

The external validity of this study was limited by the small sample size as well as other factors such as the small special education classroom and small student to teacher ratio in the classroom. Future studies should assess the effectiveness of this intervention for children with autism and other disabilities across ages and settings. Additionally, the current study did not incorporate a maintenance phase in which effects of the intervention were measured after the intervention was terminated. Since the ultimate goal of a video modeling intervention is to create a lasting behavior change, it is important that future researchers evaluate maintenance effects of video modeling and explore methods to enhance maintenance.

Overall, this study provides evidence that video modeling with the addition of an embedded rule implemented at home can result in improved transition skills in both the home and school settings. As described earlier, this study should be considered exploratory in nature due to the inclusion of only one participant as well as some of the methodological limitations described. However, this

study does provide a foundation upon which to build future research to further explore the application of video modeling as an effective and efficient intervention to improve transition skills for children with autism. This study also provides a practical intervention for parents and teachers to implement to obtain generalized effects for transition skills for children with autism.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.

Apple, A. L., Billingsley, F., & Schwartz, I. S. (2005). Effects of video modeling alone and with self-management on compliment-giving behaviors of children with high-functioning ASD. *Journal of Positive Behavior Interventions, 7*(1), 33-46.

Ayers, 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*(2), 183-196.

Cihak, D. (2011). Comparing pictorial and video modeling activity schedules during transitions for students with autism spectrum disorders. *Research in Autism Spectrum Disorders, 5,* 433-441. doi: 10.1016/j.rasd.2010.06.006

Cihak, D., Fahrenkrog, C., Ayres, K. M., & Smith, C. (2010). The use of video modeling via a video iPod and a system of least prompts to improve transitional behaviors for students with autism spectrum disorders in the general education classroom. *Journal of Positive Behavior Interventions, 12*(2), 103-115. doi: 10.1177/1098300709332346

Darden-Brunson, F., Green, A., & Goldstein, H. (2008). Video-based instruction for children with autism. In J.K. Luiselli, D.C. Russo, W. P. Christain & S.M. Wilczynski (Eds.), *Effective practices for children with autism: Educational and behavioral support interventions that work* (pp. 224-268). New York: Oxford University Press.

Delano, M. (2007). Video modeling interventions for individuals with autism. *Remedial and Special Education, 28*(1), 33-42. doi: 10.1177/07419325070280010401

Heflin, L. J., & Alaimo, D. F. (2007). *Students with autism spectrum disorders: Effective instructional practices.* Upper Saddle River, NJ: Pearson.

McCoy, K., & Hermansen, E. (2007). Video modeling for individuals with autism: A review of model types and effects. *Education and Treatment of Children, 30*(4), 183-213.

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.

Odom, S. L., Brantlinger, E., Gersten, R., Horner, R. H.,

Thompson, B., & Harris, K. H. (2005). Research in special education: Scientific methods and evidence-based practices. *Exceptional Children, 71,* 137-148.

Schreibman, L., Whalen, C., & Stahmer, A. C. (2000). The use of video priming to reduce disruptive transition behavior in children with autism. *Journal of Positive Behavior Interventions, 2*(3), 3-11. doi: 10.1177/109830070000200102

Sherer, M., Pierce, K. L., Paredes, S., Kisackv, K. L., Ingersoll, B., & Schreigman, L. (2001). Enhancing conversation skills in children with autism via video technology: Which is better, "self" or "other" as a model? *Behavior Modification, 25,* 140-158.

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*(1), 23-36. doi: 10.1177/1088357609352901

Spriggs, A. D., Knight, V., & Sherrow, L. (2014). Talking picture schedules: Embedding video models into visual activity schedules to increase independence for students with ASD. *Journal of Autism and Developmental Disorders*, Advance online publication. doi: 10.1007/s10803-014-2315-3

Sterling-Turner, H. E., & Jordan, S. S. (2007). Interventions addressing transition difficulties for individuals with autism. *Psychology in the Schools, 44,* 681-690. doi: 10.1002/pits.20257

Tullis, C. A., Cannella-Malone, H. I., & Payne, D. O. (2014). Literature review of interventions for between-task transitioning for individuals with intellectual and developmental disabilities including autism spectrum disorders. *Review Journal of Autism and Developmental Disorders, 2,* 1-12. doi: 10.1007/s40489-014-0039-0

Article

Responses to School Bomb Threats: A Survey of Stakeholders

Coady Lapierre, Gregory Ruehlen, Sherri Myers, and Samuel Fiala
Texas A&M University - Central Texas

During the winter of 2012, more than a dozen bomb threats were made against the junior high and high school of a small town in Texas which resulted in three school days cancelled, five school days ended early, and a number of responses by the school district to address the threats. A review of the literature shows there has been little formal study on how schools can respond to threats of this nature or on the effectiveness of those responses (Newman, 2005). To address this lack of knowledge, the local school district and a nearby university partnered to conduct a review of the responses that were made to the threats by surveying relevant stakeholders on their assessment of what worked and their level of approval regarding each of the responses. The results of the survey found high levels of approval for every response and generally high levels of perceived effectiveness in most groups.

Keywords: school bomb threats, school psychology, school psychologist

Late in the fall of 2012, residents of a small Texas town were anxious about numerous bomb threats in both the junior high and high school. School officials moved to find effective methods to address this rash of threats. Parents' concern grew with each new threat regarding the safety of their children and the ability of the school district to protect the students. First responders spent many hours securing the school buildings and ensuring that no bombs were actually located within any of the schools' structures. Residents of the town followed developments through media outlets and word-of-mouth. The entire community was concerned that these threats might be more than mere childish pranks. Parents lost time from work or other daily activities on days when students were either released early or not allowed to enter the school buildings at all. First responders spent hours searching school buildings for non-existent bombs. The school district spent up to $25,000 a day for each bomb threat.

The threat of violence within the school environment demands a strong and effective response. Conducting research on the effectiveness of such responses is difficult as it is challenging to predict and prepare for threats. The organizations responding to threats (i.e., police, fire, and EMS) often have all available resources committed to the response, leaving insufficient resources for research and reflection. Borum, Cornell, Modzeleski, and Jimerson (2010) note the need for researching crisis response plans to help counter confusion and unrealistic fears. In 2000, Nagy and Danitz reported that the Columbine shooting deeply affected the trust that parents had previously held about the safety of schools across the country (as cited in Borum et al., 2010). Saad (2012) noted that parents' concerns about their children's safety at school always increase immediately following incidents of school violence. Cornell and Mayer (2010) note the costs involved in interventions rather than scientific study can drive the reaction. Community apprehension over school safety may be difficult to allay due to the above factors. While school violence is of great concern, the actual number of incidents is very low with less than 1% of homicides occurring in children between the ages 5 and 18 occurring at a school (U.S. Surgeon General, 2001).

Fortunately, school violence rarely occurs, so research into the topic is sparse. School bombings are

even scarcer. The Bureau of Alcohol, Tobacco, Firearms, and Explosives reported that between 2002 and 2004, 8 people were killed and 49 were injured by explosives in the United States (Office of the Inspector General, 2005). Bombings are rare in the United States and even more atypical at schools, but bomb threats at schools may not be so uncommon. A search of governmental entities turned up no federal or state statistical records for how often schools receive bomb threats. The limited data available do seem to indicate that schools are not a common target of bomb threats in general. The Federal Bureau of Investigation reported that almost 5% of the 1,797 actual and attempted bombings in 1999 were directed at schools, but it is unknown how many bomb threats against schools were perpetrated (Newman, 2005). For comparison, one study in Finland noted that among the 69 total bomb threats studied only 4 threats were made against schools (Häkkänen, 2006). In addition to actual attacks at schools being rare, threats directed to the intended target occur in fewer than 25% of attacks. In 75% of the cases where threats were communicated before an attack, the attacker told someone not associated with the target and in 50% of these cases they told multiple people (Vossekuil, Reddy, & Fein, 2001).

This phenomenon highlights the need for school districts to work on developing a culture within the student body that encourages them to speak out about threats (Cornell, Sheras, Gregory, & Fan, 2009). This primary prevention approach has been found to be more effective than tertiary approaches in reacting to threats as they happen (Borum et al., 2010). Cornell et al. (2009) found increasing security personnel used by schools did not have an impact on the number of violent crimes reported. While it has been decades since a serious school bombing in the U.S., the largest mass murder in an American school was due to a bomb that killed 38 children and six adults and injured 58 others in Bath, Michigan during 1927 (Hoffman, 1999). So while attacks against schools are rare, they do occur, requiring school officials to respond to threats seriously.

The current research employed a cross-institutional model between a regional university and a local school district using the resources of both to evaluate responses to threats of violence as they were occurring. The authors designed the research and completed data analysis while the local school district personnel developed the interventions and consulted on the construction and dissemination of the survey. A full report of the findings was given to the local school district prior to general publication. The Institutional Review Board of Texas A&M University – Central Texas approved all research materials and procedures.

Method

Participants

Major stakeholders who were affected by the threats (and the responses to the threats) were identified for inclusion in the survey: school administrators, school staff, teachers, parents, students, community members, and first responders. Participants were recruited by email invitation (for school personnel), through the school social media (Facebook) page and main school district web page, as well as flyers and advertising in local papers.

A total of 447 responses were gathered, with one respondent removed from the data set who indicated residence outside of the school district and no children attending school within the district. Of the remaining respondents, 153 were teachers in the district, 132 were parents of a child in the district, 77 were staff, 36 were students, 22 were residents of the town without children attending the school district, 17 were school district administrators, and 9 were first responders and/or law enforcement officers who were part of response teams to the threats against the schools. No additional demographic data were collected.

Materials

A survey was constructed to measure three aspects of stakeholders' perceptions of the District's response to a series of bomb threats. The first was the perceived effectiveness of the School District's responses to the threats. The second was the acceptability of each response (i.e., their approval of the response). The last area assessed was stakeholders' view of the District's ability to communicate with the schools and community.

Each participant was asked to rate the perceived effectiveness of a number of different responses and changes the school made to standard procedures. The

following interventions were measured: students entering the school through one door, counselors' warnings of potential consequences of making a threat, installing more security cameras, employing search dogs, restricting bathroom and hall passes, scheduling make-up days, using metal detectors, and recruiting additional security personnel. After the cluster of bomb threats ended, participants were asked to rate how much they approved of each of these responses using a 5-point Likert scale ranging from 1 (*strongly disapprove*) to 5 (*strongly approve*) and how effective each of the responses was in addressing the threats made in the District using a scale that ranged from 1 (*not effective*) to 5 (*very effective*). They were also asked how effective information from the District was in stopping threats against schools. Finally, they were asked to rate how well the District kept them informed (overall, with the use of automated phone calls, and with the use of social media postings). These responses were made on a 5-point scale that ranged from 1 (*Poor*) to 5 (*Excellent*).

Results

Effectiveness and Approval of Interventions

Within-group differences in the approval and perceived effectiveness of District interventions to address a series of bomb threats were examined with a pair of repeated-measures analyses of variance (ANOVA) with a Greenhouse-Geisser correction. Post hoc pairwise comparisons with Bonferroni-adjusted alpha levels of .006 (.05/8) were then conducted to determine which interventions significantly differed from each other.

The repeated measures ANOVA examining differences in perceived effectiveness of interventions was statistically significant, $F(5.94, 2345.33) = 76.91$, $p < .001$, $\eta p^2 = .16$. Post hoc analyses (see Table 1) indicated that three of the eight interventions were rated as more effective than the other interventions in addressing the threats: hiring additional security personnel ($M = 4.19$, $SD = 0.99$), use of search dogs ($M = 4.11$, $SD = 1.02$), and installation of additional cameras ($M = 4.10$, $SD = 1.03$). As shown in Table 2, these three interventions were among the highest four interventions in perceived effectiveness for each group of stakeholder participants.

Two interventions were rated as less effective than all others: scheduling make up days for time lost to the threats ($M = 3.12$, $SD = 1.43$) and giving warnings to students ($M = 3.21$, $SD = 1.19$). Giving students warnings about the potential consequences of making bomb threats was the lowest or second lowest rated intervention for all groups surveyed except for the students themselves. Although it was the fourth highest rated intervention among students, the average rated effectiveness of being given warnings ($M = 3.31$, $SD = 1.39$) was similar in this group when compared to other groups. The use of make-up days likewise was the lowest or second lowest rated intervention for all groups surveyed except for first responders ($M = 3.88$, $SD = 1.13$) and community residents ($M = 4.18$, $SD = .91$).

The repeated measures ANOVA examining differences in approval of the District's interventions was also statistically significant, $F(5.10, 2049.51) = 82.11$, $p < .001$, $\eta p^2 = .17$. Mirroring the findings for effectiveness, post hoc analyses indicated that hiring additional security personnel ($M = 4.48$, $SD = 0.82$), installation of additional cameras ($M = 4.58$, $SD = 0.72$), and use of search dogs ($M = 4.45$, $SD = 0.84$) received higher approval ratings than the other five interventions. As shown in Table 2, these three interventions were among the highest four interventions in approval ratings for most of the groups of stakeholder participants. Adding security personnel was the fifth most approved intervention for first responders ($M = 4.44$, $SD = 0.88$), and it was the sixth most approved intervention for administrators ($M = 4.31$, $SD = 1.14$).

One intervention was given approval ratings lower than all other interventions: scheduling make up days for time lost to the threats ($M = 3.37$, $SD = 1.43$). The use of make-up days was among the least approved interventions for all groups of stakeholders.

In addition to these within-groups analyses, between-groups differences in ratings of effectiveness and approval were examined with a pair of one-way ANOVAs. The dependent variables for these analyses were the ratings of effectiveness and approval averaged across all eight interventions. Because of the small number of first-responders who provided data for these questions ($n = 9$), this group was not included in these analyses. Post hoc pairwise compa-

Table 1

Within-group Differences in Perceived Effectiveness and Approval of Interventions

	Effectiveness ($n = 396$)		Approval ($n = 403$)	
Intervention	M	SD	M	SD
Additional Security Personnel	4.19_a	.99	4.48_a	.82
Use of Search Dogs	4.11_a	1.02	4.45_a	.84
Additional Security Cameras	4.10_a	1.03	4.58_b	.72
Metal Detectors/Screening Procedures	3.89_b	1.20	4.25_c	1.04
Change Hall/Bathroom Procedures	3.78_b	1.27	3.99_{de}	1.24
Single Entry to School	3.74_b	1.28	3.90_e	1.28
Talks/Warnings with Students	3.21_c	1.19	4.17_{cd}	.99
Make-up Days for Early Release	3.12_c	1.43	3.37_f	1.43

Note. Means with differing subscripts within columns are significantly different at the $p < .006$ level.

Table 2
Effectiveness and Approval of Interventions by Stakeholders

		Effectiveness			Approval		
Stakeholder	Intervention	n	M(SD)	rank	n	M(SD)	rank
Students	Personnel	36	3.69(1.35)	1	34	3.82(1.22)	3
	Dogs	36	3.69(1.24)	2	35	3.74(1.22)	4
	Cameras	36	3.56(1.16)	3	35	4.17(.82)	1
	Metal detectors	36	3.22(1.73)	5	36	3.14(1.53)	5
	Change procedures	36	2.50(1.66)	6	36	2.08(1.36)	8
	Single entrance	35	2.49(1.36)	7	35	2.60(1.56)	6
	Warn students	36	3.31(1.39)	4	35	4.11(1.08)	2
	Make-up days	36	2.42(1.52)	8	34	2.18(1.34)	7
Parents	Personnel	132	4.36(.87)	1	132	4.70(.57)	1
	Dogs	129	4.14(1.01)	2	131	4.51(.78)	3
	Cameras	132	4.05(1.09)	3	132	4.67(.59)	2
	Metal detectors	128	4.02(1.11)	4	130	4.45(.85)	4
	Change procedures	129	3.46(1.31)	6	132	3.89(1.29)	6
	Single entrance	132	3.69(1.21)	5	133	3.78(1.29)	7
	Warn students	132	3.13(1.24)	7	130	4.32(.88)	5
	Make-up days	132	3.04(1.49)	8	132	3.37(1.48)	8
Teachers	Personnel	150	4.27(.87)	1	152	4.58(.67)	2
	Dogs	149	4.26(.87)	2	154	4.56(.76)	3
	Cameras	149	4.25(.88)	3	150	4.68(.63)	1
	Metal detectors	144	3.96(1.13)	6	149	4.26(1.01)	4
	Change procedures	152	4.13(1.00)	4	154	4.19(1.07)	5
	Single entrance	154	4.00(1.19)	5	153	4.14(1.19)	7
	Warn students	149	3.18(1.18)	8	153	4.16(1.01)	6

		N	M(SD)		N	M(SD)	
	Make-up days	153	3.24(1.38)	7	154	3.30(1.45)	8
Administrators	Personnel	16	4.31(1.01)	4	16	4.31(1.14)	6
	Dogs	16	4.31(.70)	5	16	4.81(.54)	1
	Cameras	16	4.63(.72)	2	16	4.81(.54)	2
	Metal detectors	13	4.15(1.07)	6	16	4.19(1.17)	7
	Change procedures	15	4.60(.63)	3	15	4.47(.74)	3
	Single entrance	15	4.67(.62)	1	15	4.40(.91)	4
	Warn students	15	3.13(.99)	8	16	4.38(.81)	5
	Make-up days	14	3.79(1.05)	7	15	4.13(.99)	8
Staff	Personnel	76	4.09(1.02)	2	78	4.33(.91)	3
	Dogs	77	3.99(1.12)	4	78	4.41(.87)	2
	Cameras	77	4.16(1.11)	1	78	4.51(92)	1
	Metal detectors	71	3.93(1.13)	5	78	4.26(1.02)	5
	Change procedures	79	4.04(1.09)	3	78	4.28(1.01)	4
	Single entrance	76	3.87(1.21)	6	79	4.10(1.09)	6
	Warn students	74	3.28(1.14)	7	77	3.99(1.13)	7
	Make-up days	77	3.09(1.36)	8	78	3.42(1.33)	8
First responders	Personnel	9	4.22(1.09)	3	9	4.44(1.22)	5
	Dogs	9	4.22(1.39)	3	9	4.78(.67)	1
	Cameras	9	4.56(.73)	1	9	4.67(.50)	3
	Metal detectors	9	3.67(1.32)	6	9	4.33(.87)	7
	Change procedures	8	4.38(.74)	2	8	4.75(.46)	2
	Single entrance	9	3.67(1.22)	7	9	3.67(1.41)	8
	Warn students	8	3.50(1.07)	8	9	4.67(.50)	3
	Make-up days	8	3.88(1.13)	5	9	4.44(.73)	5
Community residents	Personnel	36	3.69(1.35)	1	22	4.55(.80)	1
	Dogs	36	3.69(1.24)	2	22	4.45(.74)	3
	Cameras	36	3.56(1.16)	3	21	4.48(.87)	2
	Metal detectors	36	3.22(1.73)	5	22	4.45(.74)	3
	Change procedures	36	2.50(1.66)	6	21	4.24(10.7)	6
	Single entrance	35	2.49(1.36)	7	22	4.09(1.19)	7
	Warn students	36	3.31(1.39)	4	22	4.05(1.0)	8
	Make-up days	36	2.42(1.52)	8	22	4.23(1.07)	6

risons with Bonferroni-adjusted alpha levels of .008 (.05/6) were then conducted to determine which interventions significantly differed from each other.

The ANOVA examining group differences in perceived effectiveness of the interventions was statistically significant, $F(5, 425) = 7.78$, $p < .001$, $\eta p^2 = .08$. Post hoc pairwise comparisons indicated that students rated the interventions as less effective than did any other group. Similarly, the ANOVA examining group differences in approval of the interventions was statistically significant, $F(5, 434) = 16.11$, $p < .001$, $\eta p^2 = .16$. Again, students reported lower approval of the interventions than any other group. There were no other statistically significant group differences (See Table 3).

Communication

The ANOVA examining differences among groups (sans first responders) in perceptions of how effective the District's communication was in reducing or stopping threats to schools was statistically significant, $F(5, 430) = 2.35$, $p < .05$, $\eta p^2 = .03$. Post hoc pairwise comparisons (with an adjusted alpha) indicated that teachers ($M = 2.97$, $SD = 1.18$) were slightly more positive about the impact of the District's communication on the cessation of bomb threats than were parents ($M = 2.55$, $SD = 1.18$). How-

Table 3

Between-groups Differences in Perceived Effectiveness and Approval of Interventions

Stakeholder group	Effectiveness ($n = 431$)		Approval ($n = 440$)	
	M	*SD*	*M*	*SD*
Students	3.11_a	.91	3.23_a	.89
Parents	3.73_b	.83	4.21_b	.58
Teachers	3.92_b	.71	4.24_b	.62
Administrators	4.19_b	.47	4.44_b	.60
Staff	3.81_b	.81	4.17_b	.72
Community residents	4.06_b	.65	4.32_b	.65

Note. Means with differing subscripts within columns are significantly different at the $p < .006$ level.

ever, no group's average for this item was above the scale's midpoint. The ANOVA examining group differences in how well informed each participant reported being was not statistically significant, $F(5, 429) = 1.99$, $p = .079$, $\eta p^2 = .02$. Although participants tended to rate the District's communication as somewhat ineffective in reducing or stopping threats ($M = 2.80$, $SD = 1.20$), they tended to positively rate the District's efforts at keeping them informed ($M = 3.46$, $SD = 1.30$).

A stepwise multiple regression analysis was conducted to examine the relationship between different avenues of communication (i.e., phone calls and social media postings) and satisfaction with District communication efforts. The two-predictor model accounted for 42.25% of the variance in participants' reports of how well informed they felt, $F(2, 403) = 147.40$, $p < .001$, $R^2 = .42$. Satisfaction with communication via phone was most strongly related to feeling well informed by the District ($\beta = .58$, $p < .001$) and accounted for 40.62% of the variability in responses about being informed. Satisfaction with social media communiqués from the District was also a significant predictor ($\beta = .14$, $p < .001$) but only contributed an additional 1.63% of explained variance. However, a paired-groups t-test comparing ratings of the usefulness of phone calls ($M = 3.62$, $SD = 1.29$) vs. social media postings ($M = 3.53$, $SD = 1.32$)

suggests that participants found them equally helpful for staying informed, $t(412) = 1.29$, $p = .199$.

Discussion

In general, the district received ratings above the scale mid-point for all of their responses and communication during the period of bomb threats. Additional security personnel, employing search dogs, and increasing security cameras were highly rated interventions, both in terms of perceived effectiveness and in regards to approval. The popularity of these responses may be due to their shared focus on their potential for identifying a threat without adding undue burden on students, faculty, and staff. The other District response that focused on threat identification (i.e., increased use of metal detectors) may not have been rated as highly as these other interventions because of the increased inconvenience associated with its implementation. Other District responses that were rated as moderately effective (e.g., changing procedures for hall passes and limiting schools to one entrance) were also likely seen as an encumbrance. Not surprisingly, the least popular intervention (i.e., scheduling make up days) was likely the most inconvenient.

Future research may need to address the other response that was viewed as relatively ineffective: warning students about the consequences of making

bomb threats. This strategy ranked near the bottom of the list for most participants in terms of effectiveness, but it was the fourth most effective intervention in the eyes of the students themselves. Moreover, there was only one other District response that they approved of more than this one.

This disconnect between students and other stakeholders is reflected more clearly when looking at overall ratings of effectiveness and approval of the District's responses. The group that was impacted the most by these interventions, the students, are the ones who rated the changes the lowest.

Although the District's communication was not viewed as an effective tool in the reduction of bomb threats, relevant stakeholders did report feeling well informed by the District. Both phone messages and social media postings from the District were viewed positively, and satisfaction with phone messages in particular was predictive of how well informed the stakeholders reported being. Future research may reveal benefits of this level of communication other than directly reducing/stopping the threats (e.g., decreased anxiety about the threats).

The information presented here will help school personnel be more knowledgeable about community perceptions involving responses to threats to schools. This project also shows a model of cooperation between institutions that can prove mutually beneficial. School districts should be encouraged to partner with local universities to plan for external review of emergency responses. Districts will gain valuable feedback to improve their systems while universities gain valuable opportunities to study an important and underpopulated area of study.

References

Borum, R., Cornell, D., Modzeleski, W., & Jimerson, S. (2010). What can be done about school shootings? A review of the evidence. *Educational Researcher, 39*(1), 27-37. doi:10.3102/0013189X0935762

Cornell, D., & Mayer, M. (2010). Why do school order and safety matter? *Educational Researcher, 39*(1), 69-78. Retrieved from http://www.jstor.org/stable/27764549

Cornell, D., Sheras, P., Gregory, A., & Fan, X. (2009). A retrospective study of school safety conditions in high schools using the Virginia threat assessment guidelines versus alternative approaches. *School Psychology Quarterly, 24*(2), 119-129. doi:10.1037/a0016182

Häkkänen, H. (2006). Finnish bomb threats: Offence and offender characteristics. *International Journal of Police Science and Management, 8*(1), 1–8. doi:10.1350/ijps.2006.8.1.1

Hoffman, K. B. (1999, April 23). Bath tragedy still among worst: 38 students, 5 adults killed in 1927 bombing. *Ludington Daily News*, p. 7.

Newman, G. R. (2005). *Bomb threats in schools.* (Report No. 32). U.S. Department of Justice. Retrieved from www.cops.usdoj.gov/.../e061120371_POP_Bomb-ThreatsinSchools.pdf

Office of the Inspector General. (2005). *Review of the Bureau of Alcohol, Tobacco, Firearms and Explosives' implementation of the safe explosives act.* Retrieved from http://www.justice.gov/oig/reports/ATF/e0505/app1.htm

Saad, L. (2012, December 28). *Parents' fear for children's safety at school rises slightly: In U.S., 52% say it is likely school shooting could occur in their area.* Retrieved from http://www.gallup.com/poll/159584/parents-fear-children-safety-school-rises-slightly.aspx

U.S. Surgeon General. (2001), *Youth Violence: A Report of the Surgeon General.* Rockville, MD: U.S. Department of Health and Human Services.

Vossekuil, B., Reddy, M., & Fein, R. (2001). The Secret Service safe school initiative. *Education Digest, 66*(6), 4-7.

Article

The Baylor Revision of the
Motivation to Read Survey (B-MRS)

Marley W. Watkins and Larry J. Browning
Baylor University

Although reading motivation has been recognized as important by both teachers and researchers, scales to measure reading motivation have not been well validated. The structural validity of one promising measure of reading motivation, the Reading Survey (MRS) portion of the Motivation to Read Profile, was investigated in the current study with initial ($N = 933$), calibration ($N = 545$), and normative ($N = 2,146$) samples. Data from the initial sample revealed the anticipated two factors of reading self-concept and value of reading, but identified three problematic items. Those items were replaced and the resulting Baylor revision of the MRS (B-MRS) was administered to the calibration sample. Exploratory factor analysis of the calibration sample data revealed the anticipated two factors with no problematic items. Confirmatory factor analysis was then applied to B-MRS data from the normative sample. Using multilevel methods because students were nested in classrooms, a two-factor theoretical structure was found to fit within students whereas a one-factor model best fit between classrooms. Girls tended to have more positive reading self-concepts and valued reading at higher levels than boys while both reading self-concept and value of reading scores decreased as grade level increased. Norms were described and use of the B-MRS by practitioners and researchers was encouraged.

Keywords: factor analysis, validity, reading motivation, reading

The importance of reading to individuals and society cannot be overstated. Consequently, there is an obvious need to teach all children to read at proficient levels. Unfortunately, this need is not being effectively met. Data from the National Assessment of Educational Progress (NCES, 2013) indicates that 32% of tested 4th grade students and 22% of tested 8th grade students were Below Basic in reading skills. Below Basic means that these students did not demonstrate even partial mastery of the reading skills needed for proficient work at their grade level (NCES, 2013).

Reading skill and its development has been a major research interest in education and psychology for decades (Adams, 1990). Historically, most research focused on the cognitive aspects of reading such as phonemic awareness, word reading, fluency, vocabulary, and comprehension (Gough & Tunmer, 1986; LaBerge & Samuels, 1974; Perfetti & Stafura, 2014; Rayner, Foorman, Perfetti, Pesetsky, & Seidenberg,

2001; Snow, Burns, & Griffin, 1998). Stanovich (1986) noted that nearly every cognitive task that comprises the act of reading has been investigated, and more recent research has provided considerable guidance for fostering development of skilled readers (Hairrell, Rupley, & Simmons, 2011; Hattie, 2012; Hulme & Snowling, 2011; Marulis & Neuman, 2013; Piasta & Wagner, 2010; Slavin, Lake, Chambers, Cheung, & Davis, 2009; Snowling & Hulme, 2011).

Although important to teachers (O'Flahavan, Gambrell, Guthrie, Stahl, & Alvermann, 1992), aff-

ective facilities have only recently been recognized as potentially important contributors to reading proficiency (Afflerbach & Cho, 2011; Graham & Weiner, 1996; Hidi & Harackiewicz, 2000). Reading motivation in particular has garnered a substantial amount of attention as it applies to student learning (Graham & Weiner, 2012). For example, an early quantitative synthesis of the research on motivation and achievement found that around 11% of the variance in achievement was accounted for by motivation (Uguroglu & Walberg, 1979). More recently, a meta-analysis of 69 data sets involving more than 125,000 students found that verbal achievement and verbal self-concept were related at $r = .49$ (Möller, Pohlmann, Köller, & Marsh, 2009). These results seem to be consonant with commonsense views: It is intuitively pleasing to assume that students who read well do so partly because they are motivated to read, and those students who do not read well struggle partly because they are not motivated (Stanovich, 1986).

Over time, researchers have examined both affective and cognitive variables and considered the potential for interaction and synergy between cognitive *skill* and motivational *will* (Linnenbrink & Pintrich, 2002; Paris & Oka, 1986). In fact, considerable evidence has accumulated to suggest that affective and cognitive variables are reciprocally related and mutually reinforcing (Chamorro-Premuzic, Harlaar, Greven, & Plomin, 2010; Marsh, Xu, & Martin, 2012; Morgan & Fuchs, 2007; Retelsdorf, Köller, & Möller, 2014). Simultaneously, motivation theories (e.g., expectancy-value theory, self-determination theory, attribution theory, goal theory, etc.) were developed to explain the accumulating empirical findings and guide future research (Graham & Weiner, 2012; Pintrich & Schunk, 2002). These theories posited an array of constructs to explain motivated reading behavior (Anderman, Gray, & Chang, 2013; Guthrie & Coddington, 2009; Murphy & Alexander, 2000; Schiefele, Schaffner, Möller, & Wigfield, 2012), such as intrinsic and extrinsic motivation, perceived autonomy, self-concept, self-efficacy, task mastery goals, performance goals, prosocial goals, compliance goals, value, and autonomous motivation (Conradi, Jang, & McKenna, 2014).

As with all theories, those concerning reading motivation "can be divided into two parts: one that specifies relationships between theoretical constructs and another that describes relationships between constructs and measures" (Edwards & Bagozzi, 2000, p. 155). Most of the research on reading motivation has dealt with the relationships between theoretical constructs (i.e., reading motivation and reading achievement) and relatively little attention has focused on the relationship between constructs and measures. That is unfortunate because a robust construct-measure relationship allows an unambiguous mapping of theoretical constructs onto empirical measures and is, in effect, an auxiliary theory (Edwards & Bagozzi, 2000). In educational and psychological measurement, the construct-measure relationship is often recognized under the rubric of structural validity. That is, whether the structure of scores generated by a measure reflects the theoretical structure of the construct (Messick, 1995). Strong structural validity evidence facilitates both research and practice (Kane, 2013) and should precede research on the relationships between constructs (Meehl, 1990).

Recognizing a need for valid measures of reading motivation (Schunk, 2000), researchers have developed a number of scales designed to measure reading motivation, but most are distinguished by "poor construction and limited validation" (Fulmer & Frijters, 2009, p. 226). Two scales that have received considerable attention are the Motivations for Reading Questionnaire (MRQ; Wigfield, Guthrie, & McGough, 1996) and the Survey portion of the Motivation to Read Profile (MRS; Gambrell, Palmer, Codling, & Mazzoni, 1996). The MRQ is a 54-item group administered scale with a 4-point response format that purports to measure 11 aspects of reading motivation. Although the MRQ has frequently been applied in reading research (e.g., Klauda & Wigfield, 2012), an extensive analysis of its structural validity concluded that the MRQ should not be used (Watkins & Coffey, 2004, p. 117).

The MRS is a 20-item group administered survey with a 4-point response format that measures two aspects of reading motivation: self-concept as a reader and value of reading. An individually administered interview was also included in the Motivation to Read Profile but will not be considered further because it uses an open-ended question format and was not designed to be scored. Based on expectancy-value theory (Eccles, 1983), the items in the reading self-concept scale were designed to assess students' expectations of success in reading and the items in

the value of reading scale were designed to measure the value students ascribe to reading. In essence, "Can I do it? Do I want it?" (Graham & Weiner, 2012, p. 372).

Structural validity for the MRS was supported by exploratory factor analysis (EFA) using unweighted least squares extraction and varimax rotation on data from 330 third- and fifth-grade students in 27 classrooms in four schools from two school districts in an Eastern state (Gambrell et al., 1996). Subsequently, the MRS has been applied in reading research (Applegate & Applegate, 2011; Gambrell, Hughes, Calvert, Malloy, & Igo, 2011; Marinak, 2013; Marinak & Gambrell, 2008, 2010; Quirk, Schwanenfugel, & Webb, 2009) and has twice been revised for use with adolescents (Kelley & Decker, 2009; Pitcher et al., 2007). Unfortunately, research on the MRS has not attended to its psychometric properties nor its structural validity.

The MRS was recently revised by Malloy, Marinak, Gambrell, and Mazzoni (2013) to modernize and update its content. The revised MRS retained seven of the original items, replaced one item, and modified 12 items. Modifications tended to be minor. For example, "almost never" replaced "not very often" as one response option and "My friends think reading is" replaced the original stem of "My best friends think reading is." The revised MRS was administered to 281 students in three schools in Virginia, South Carolina, and Pennsylvania. Alpha coefficients for the reading self-concept and value of reading scales were .81 and .85, respectively. It is not clear if the structural validity of the revised scale was evaluated because almost no methodological details were provided. For example, Malloy et al. (2013) simply reported that "a nonparametric analysis was used to determine validity using a root mean square error of approximation (RMSEA)...estimate of .089 was revealed" (p. 275). An RMSEA value of .089 would reflect a less-than-adequate overall fit of the model to the data, individual parameter estimates were evidently not reviewed, and there has been no other research on this revision of the MRS.

Critically, the structural validity investigations of the original MRS (Gambrell et al., 1996) and revised MRS (Malloy et al., 2013) were also methodologically flawed. For example, the four-option item responses constitute ordered categories rather than continuous values. In such cases, polychoric correla-

tions should be submitted to factor analysis rather than Pearson product moment correlations (Flora, LaBrish, & Chalmers, 2012). Further, analyzing individual student data for classes of students violates the fundamental assumption of independence with an attendant risk of biased parameter estimates (De Naeghel & Van Keer, 2013; Muthén, 1994). Also, sub-dimensions of reading motivation (i.e., reading self-concept and reading value) are likely to be related to some extent and forcing them to be orthogonal with a varimax rotation as done by Gambrell et al. (1996) may have resulted in a distorted factor solution (Gorsuch, 1997). Finally, no other details of the factor analyses were reported by Gambrell et al. (1996) or Malloy et al. (2013), which does not allow for informed review and replication (Ford, MacCallum, & Tait, 1986).

Given these lacunae, the current study was designed to analyze the structural validity of the original MRS and use that evidence to revise the MRS and collect validity evidence on the new revision. The resulting validity evidence will allow practitioners to better identify, implement, and evaluate interventions to improve reading motivation and achievement (Guthrie, 2011) and will provide researchers with a measure for use in future research.

Original MRS Study

Method

Instrument. The Reading Survey portion of the Motivation to Read Profile is a 20-item (each with four response options) scale for students in grades 2-6. As described by Gambrell et al. (1996), it is a "public-domain instrument" (p. 519) with 10 reading self-concept items "designed to elicit information about students' self-perceived competence in reading and self-perceived performance relative to peers" and 10 value of reading items "designed to elicit information about the value students place on reading tasks and activities" (p. 522).

The first validation study by Gambrell et al. (1996) included 330 third- and fifth-grade students in 27 classrooms in four schools from two school districts in an Eastern state. Gambrell et al. (1996) found that internal consistency reliability was .75 and .82 for the self-concept and value scales, respectively. An alpha coefficient of .89 was subsequently compu-

ted for the total MRS score with a group of 219 students in seven grade 3-5 classrooms (Gambrell et al., 2011).

Participants. A total of 933 students (48.8% male) in grades 1 through 5 (5.5% in first grade with 55% male, 7.5% in second grade with 54% male, 30% in third grade with 48% male, 32% in fourth grade with 45% male, and 25% in fifth grade with 52% male) from Arizona ($n = 340$), Maryland ($n = 333$), and Pennsylvania ($n = 260$) completed all 20 MRS items. No further demographic information was collected on individual students to protect participants' confidentiality.

Procedures. Data were obtained from elementary schools in Arizona ($n = 1$), Maryland ($n = 2$), and Pennsylvania ($n = 1$) secondary to other research projects or local program evaluations (Neuhard, 2004; Runge, 1998; Young, 2000). Students were enrolled in 42 separate classrooms with an average class size of 22.2 students. The Pennsylvania school was located in a rural area, fewer than 1% of its students were minority, and around 35% of its students received free or reduced lunch. The Maryland schools were in suburban areas, around 28% of their students were minority and 43% received free or reduced lunch. The Arizona school was also in a suburban area. Around 25% of its students were minority and around 30% received free or reduced lunch. Directions and test items were read aloud to students by researchers or teachers following the instructions provided by Gambrell et al. (1996). Unit weighted reading self-concept and value of reading scores were computed as per Gambrell et al. (1996).

Analyses. Given the well-developed theoretical expectation of two MRS factors, confirmatory factor analysis (CFA) was implemented with Mplus version 7.2 (Muthén & Muthén, 2014). Based on the ordered categorical data, polychoric correlations and the WLSMV estimator were selected (Lei & Wu, 2012). Overall model fit was evaluated with the comparative fit index (CFI) and the root mean square error of approximation (RMSEA). Criteria for adequate model fit were CFI $\geq$ .90 and RMSEA $\leq$ .08 whereas good model fit required CFI $\geq$ 0.95 and RMSEA $\leq$ 0.06 (Hu & Bentler, 1999). Intraclass correlations for items ranged from .03 to .08 with a median of .06, indicating that non-independence of student data should be considered in the analyses (Muthén, 1997).

Results

As suggested by Hox (1995), a baseline model was established by comparing one- and two-factor models without regard for nested data. The two-factor model was clearly superior to the one-factor model (CFI of .95 vs. .83 and RMSEA of .07 vs. .14, respectively) although its overall fit was only adequate. Next, the non-independence of student data was taken into account with the Mplus cluster procedure in a two-factor model, which exhibited good fit to the data (CFI of .96 and RMSEA = .05). Thus, the two-factor structure of the Reading Survey portion of the Motivation to Read Profile was supported. The two factors correlated at .60 and exhibited alpha coefficients of .82 (95% CI [.78, .85]) for the reading self-concept factor and .84 (95% CI [.80, .87]) for the value of reading factor.

Regardless of overall model fit, inspection of the standardized parameter estimates revealed two problems with the two-factor model: one reading self-concept item (#11) and one value of reading item (#18) were weakly related to their respective factors (.12 and .21, respectively) in comparison to the remainder of the items (Md = .70). Additionally, several students spontaneously wrote critical comments on protocols about the stem of item 17 (*When I am in a group talking about stories*) indicating that talking about stories in a group was only for primary grade students. Interestingly, the lack of ecological validity of item #17 may have been noticed in prior studies because its stem was revised in both attempts to create an adolescent version of the MRS (i.e., Kelley & Decker, 2009; Pitcher et al., 2007). Altogether, then, three of the 20 items on the Survey portion of the Motivation to Read Profile were problematic and required revision or replacement.

Results for the reading self-concept and value of reading scales across grade level and sex are illustrated in Figure 1. Regression analyses were conducted using clustered robust standard errors within Stata 13 to adjust for non-independence of the data. For reading self-concept, grade was a statistically significant predictor ($t = -2.47$, $df = 4$, $p = .018$, $R^2 = .02$), but neither sex nor the grade by sex interaction were significant predictors ($p > .05$). In contrast, both grade ($t = -4.10$, $df = 4$, $p < .001$, $R^2 = .06$) and sex ($t = 2.61$, $df = 1$, $p = .013$, $R^2 = .02$) were significant
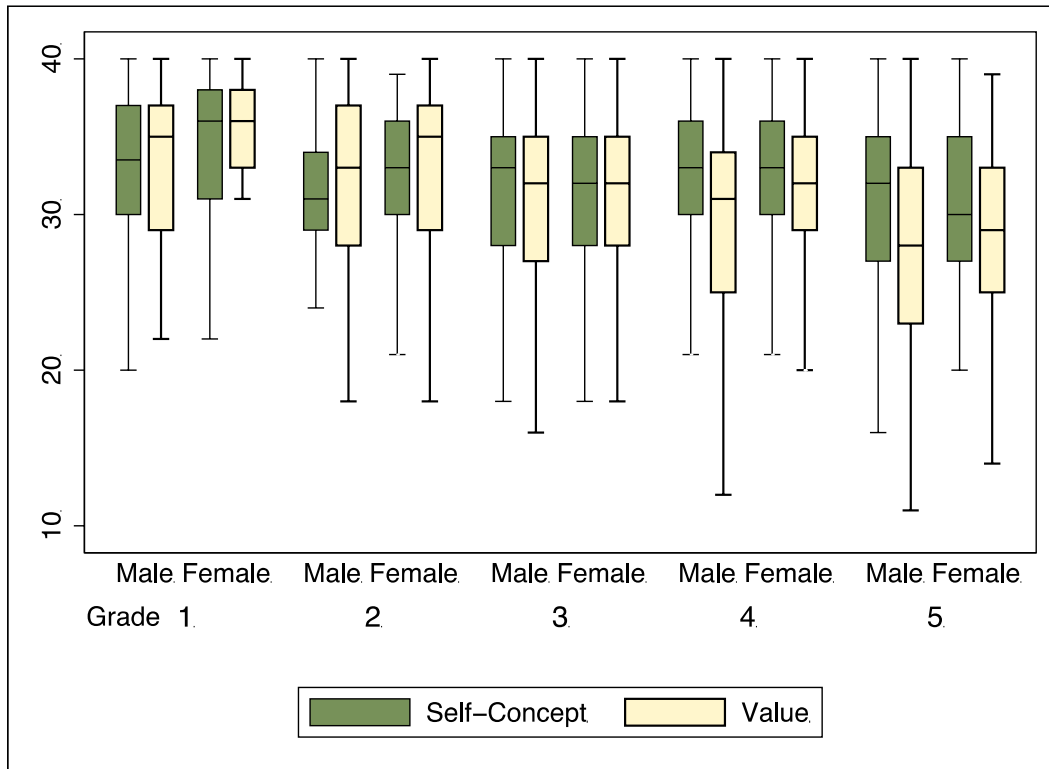
Figure 1. Reading Self-Concept and Value of Reading Scores on the Motivation to Read Survey (MRS) for 933 Male and Female Students in Grades 1 Through 5.

predictors for the value of reading scale, but the grade by sex interaction was not a significant predictor ($p > .05$). The correlation between grade and reading self-concept was -0.10 and between grade and the value of reading was -0.23 indicating that both types of reading motivation declined as grade level increased. The correlation between sex and the value of reading was 0.13, indicating that girls tended to value reading more than boys.

Calibration Study

Method

Instrument. Seventeen items of the Reading Survey portion of the Motivation to Read Profile were retained. Several alternatives were created for items 11, 17, and 18 and qualitatively reviewed by a reading expert and a psychometrician. Following pilot tests with small samples of volunteer students, new items 11, 17, and 18 were selected based on psychometric performance and incorporated into the Baylor revision of the MRS (B-MRS). The B-MRS

scale as well as standardized administration instructions and score sheets can be freely downloaded from http://edpsychassociates.com.

Participants. A total of 545 students (49.5% male) in grades 1 through 8 (10% in first grade with 48% male, 17% in second grade with 48% male, 17% in third grade with 45% male, 15% in fourth grade with 56% male, 15% in fifth grade with 51% male, 3% in sixth grade with 66% male, 18% in seventh grade with 45% male, and 5% in eighth grade with 56% male) served as participants. No further demographic information was collected on individual students to protect participants' confidentiality.

Procedures. Data were obtained from a school in rural Pennsylvania secondary to local program evaluation activities. No further identifying information about students or teachers was made available. However, the school district enrolled around 9% minority students and offered free/reduced lunch to approximately 57% of its students. Directions and test items were read aloud to students by teachers following standardized instructions.

Analyses. Given the scale revision and attendant

uncertainty about its structure, an EFA using polychoric correlations was applied with the *psych* package within the **R** program (R Development Core Team, 2014). Following the best practice EFA guidelines (Ford et al., 1986; Gorsuch, 1997), the number of factors to retain for rotation was determined by parallel analysis and minimal average partials (MAP) criteria, principal axis extraction with oblimin rotation were specified, and pattern coefficients ≥ .32 were predetermined to be salient.

Results

Both parallel analysis and MAP criteria indicated that two factors should be extracted. Three factors were extracted as a deliberate over-extraction strategy. As expected, the resulting third factor was inadequate, being loaded by six items of which four were complex, leaving only two items to uniquely identify the third factor. In contrast, the two-factor solution clearly identified ten items for each factor (see Table 1). Internal consistency reliability was strong, .84 (95% CI [.80, .87]) for the reading self-concept factor and .87 (95% CI [.84, .90]) for the value of reading factor.

Results for the reading self-concept and value of reading scales across grade level and sex are illustrated in Figure 2. There were too few students in several cells for a valid test across grade levels, but the correlation between grade and reading self-concept was -0.07 and between grade and the value of reading was -0.36. Thus, reading motivation decreased as grade level increased. Additionally, males and females did not significantly differ on the reading self-concept scale ($t = 0.59$, $df = 531$, $p = .56$, $R^2 < .01$) but were statistically different in favor of females on the value of reading scale ($t = 4.29$, $df = 531$, $p < .001$, $R^2 = .04$).

Normative Study

Method

Instrument. Given the clear factor structure found in the calibration study, B-MRS data were collected and analyzed in a validation study.

Participants. A total of 2,136 Texas students (52% male) in grades 2 through 6 served as participants. By grade level, there were 301 students (53% male) in grade 2, 269 students (54% male) in grade 3, 372 students (49% male) in grade 4, 588 students (51% male) in grade 5, and 606 students (53% male) in grade 6. No further demographic information on individual students was collected to protect participants' confidentiality.

Procedures. The names and work addresses of 1,000 randomly selected Texas teachers in grades 2-6 were purchased from a commercial marketing firm. Those 1,000 teachers were solicited via U.S. Mail to anonymously collect and share B-MRS data from their classrooms. Each solicitation letter contained a classroom supply of B-MRS forms and standardized instructions for data collection. Responses were received from 88 teachers who provided B-MRS data for 2,371 students in grades 1-8. The data of 2,136 students in 83 grade 2-6 classrooms who completed all 20 B-MRS items were retained for the validation study.

Analyses. Given theory and results of the calibration study, CFA was implemented with Mplus version 7.2 (Muthén & Muthén, 2014). Based on the ordered categorical data, polychoric correlations and the WLSMV estimator were selected (Lei & Wu, 2012). Overall model fit was evaluated with the CFI and RMSEA. Criteria for adequate model fit were CFI ≥ .90 and RMSEA ≤ .08 whereas good model fit was set at CFI ≥ 0.95 and RMSEA ≤ 0.06 (Hu & Bentler, 1999). Intraclass correlations for items ranged from .05 to .21 with a median of .09 indicating that non-independence of student data should be considered in the analyses (Muthén, 1997).

Results

A baseline model was established by comparing one- and two-factor within-student models without regard for nested data (Hox, 1995). The two-factor model was clearly superior to the one-factor model (see Table 2) although its overall fit was only adequate. Consequently, multilevel models with two within-student and one and two between-classroom factors were analyzed. All multilevel models exhibited good fit to the data but four residual item variances were negative in the model with two between-classroom factors, making that model inadmissible. Fewer factors are often found at the between level of multilevel models (Brown, 2013) so this result was not unusual. Thus, the two within-student and one

Table 1

Pattern Coefficients From An Exploratory Factor Analysis of the Baylor Revision of the Reading Survey (B-MRS) Among A Calibration Sample of 545 Students in Grades 1 Through 8

Item	Self-Concept	Value of Reading	h^2
1. My friends think I	**.61**	.23	.57
2. Reading a book	.26	**.62**	.61
3. I read	**.57**	.07	.36
4. My best friends think reading	-.06	**.59**	.32
5. Don't know a word	**.59**	-.02	.34
6. Tell friends about books	.13	**.55**	.38
7. Understand silent reading	**.67**	-.11	.39
8. People who read	-.08	**.75**	.51
9. I am	**.77**	.13	.70
10. I think libraries	-.12	**.85**	.64
11. I have trouble with reading	**.75**	-.19	.47
12. Knowing how to read	.14	**.46**	.29
13. Teacher question about reading	**.46**	.21	.36
14. I think reading	.10	**.80**	.72
15. Reading is	**.77**	-.04	.56
16. When I grow up	.02	**.69**	.49
17. Talk about reading assignments	**.58**	.08	.39
18. Want teacher to spend time on reading	-.03	**.66**	.43
19. When I read aloud	**.51**	.22	.41
20. When receive book as a present	.06	**.74**	.59

Note. Salient coefficients ($\geq$ .32) in bold. h^2 = communality. Item stems abbreviated.

between-classroom factor model was the best fit to the data and is illustrated in Figure 4. Alpha coefficients for the reading self-concept and value of reading scales were both .87 with 95% CIs [.85, .89].

Reading self-concept and value of reading scores across grade level and sex are illustrated in Figure 3. Regression analyses were conducted using clustered robust standard errors within Stata 13 to adjust for non-independence of the data. For the reading self-concept scale, neither grade, sex, nor the grade by sex interaction were significant predictors ($p > .05$) with all three predictors combined accounting for less than 2% of the variance in reading self-concept. In contrast, both grade ($t = -5.55$, $df = 4$, $p < .001$, $R^2 = .11$) and sex ($t = 1.96$, $df = 1$, $p = .054$, $R^2 = .03$), but not the interaction of grade and sex ($p > .05$), were significant predictors of the value of reading

scale. The correlation between grade and the value of reading was -0.32 indicating that the perceived value of reading declined as grade level increased. The correlation between sex and the value of reading was 0.16, indicating that girls tended to place higher value on reading than boys.

Norms. The original MRS lacks a representative normative sample, which is a "minimal requirement for using a test for diagnostic purposes" (Bear, Minke, & Manning, 2002, p. 423). Without a standardized, normative comparison, MRS scores are essentially uninterpretable for clinical use and do not allow advances in research on reading motivation. The lack of norms was seen as a specific weakness of the MRS by Fulmer and Frijters (2009), and other researchers have recommended the development of norms for the MRS (Kelley & Decker, 2009).
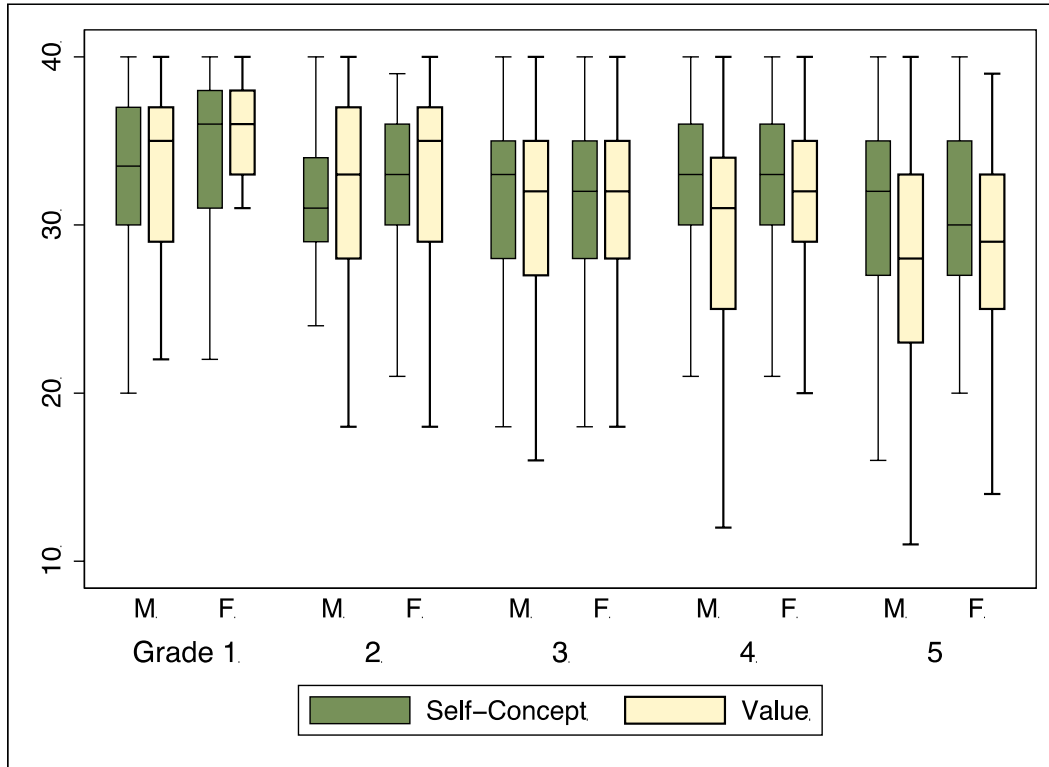
Figure 2. Reading Self-Concept and Value of Reading Scores on the Baylor Revision of the Motivation to Read Survey (B-MRS) for the Calibration Sample of 545 Male and Female Students in Grades 1 Through 5.

Table 2

Fit of Statistical Models to the Baylor Revision of the Reading Survey (B-MRS) Among A Texas Normative Sample of 2,136 Students in Grades 2 Through 6

Model	X^2	df	CFI	RMSEA
One Within-Student factor	4058.5	170	.831	.103
Two Within-Student factors	1666.9	169	.957	.064
Two Within-Student factors & One Between-Classroom factor	1262.6	339	.955	.036
Two Within-Student factors & Two Between-Classroom factors	1281.5	338	.954	.036

Note. CFI is the comparative fit index and RMSEA is the root mean square error of approximation.
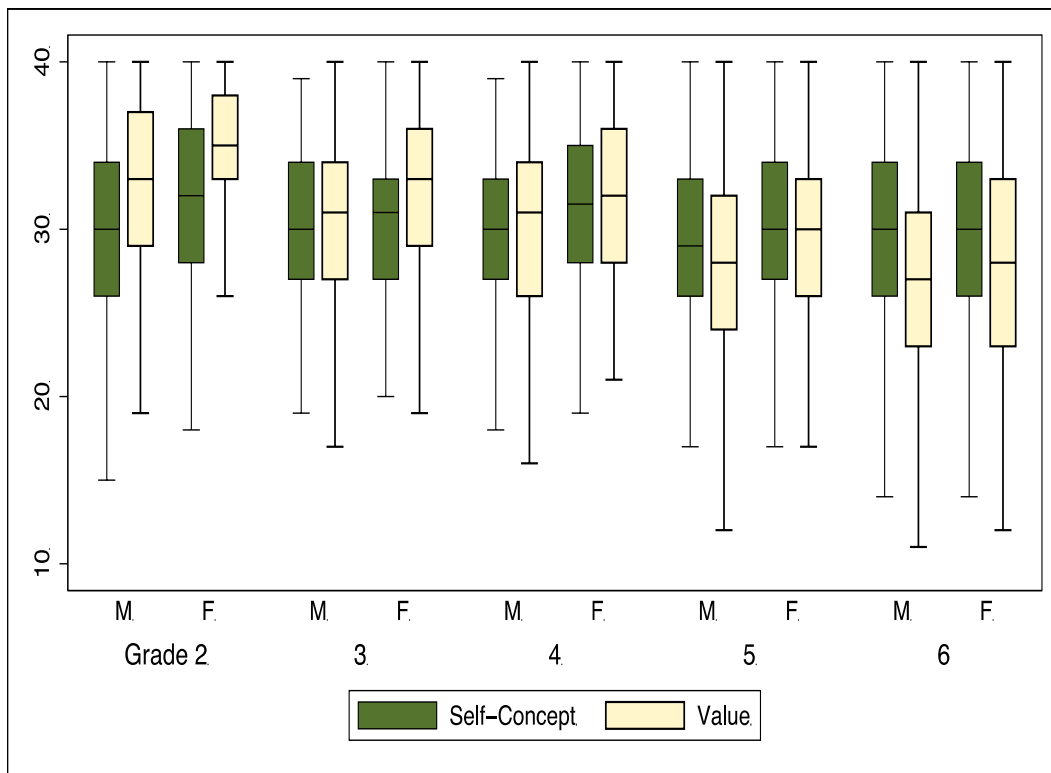
Figure 3. Reading Self-Concept and Value of Reading Scores on the Baylor Revision of the Motivation to Read Survey (B-MRS) for the Normative Sample of 2,136 Male and Female Texas Students in Grades 2 Through 6.
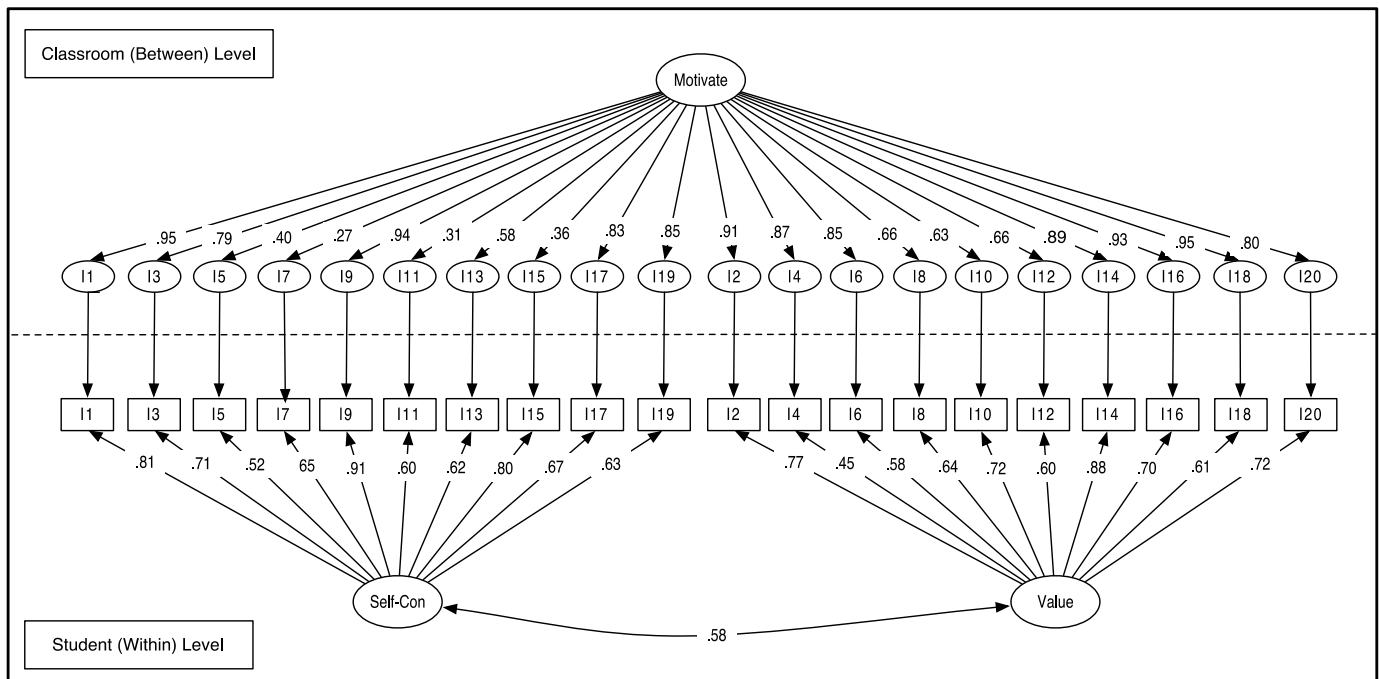

Figure 4. Multilevel structural model of the validation sample of 2,136 Texas students on the 20 items that comprise the Baylor Revision of the Motivation to Read Survey (B-MRS). Motivate is General Reading Motivation, Self-Con is Reading Self-Concept, and Value is Value of Reading.

Texas norms for the B-MRS were developed with the validation sample based on unit weighted raw scores (Bobko, Roth, & Buster, 2007; Wainer, 1976) and percentiles for their simplicity (Salvia, Ysseldyke, & Bolt, 2010). Given the unequal number of males and females at each grade level, students' scores were weighted to achieve an overall sample size of 1,500 with 150 boys and 150 girls at each of the five grade levels. Separate norms tables were created for each grade level due to the major influence of grade level, especially on the value of reading scale. Norms tables as well as the B-MRS scale, standardized administration instructions, and score sheets can be freely downloaded from http://edpsychassociates.com.

Discussion

Although reading motivation has been recognized as important by both teachers and researchers, scales to measure reading motivation have, unfortunately, been of "poor construction and limited validation" (Fulmer & Frijters, 2009, p. 226). The structural validity of one promising measure of reading motivation, the Reading Survey (MRS) portion of the Motivation to Read Profile was investigated in the current study. Based on that initial investigation, the MRS was revised and its psychometric properties and structural validity examined in calibration and normative samples. The revised scale was found to measure two related ($r = .58$) reading motivation factors within students (reading self-concept and value of reading), both with good reliability ($\alpha = .87$), and one factor between classrooms. Thus, it appeared that students differentiated the value of reading from reading self-concept, and teachers unitarily influenced both facets of reading motivation (De Naeghel & Van Keer, 2013).

Female students in grades 2-6 exhibited more positive reading self-concept and value of reading scores than did male students. However, the male-female differences were small for both reading self-concept (less than 1% of variance) and value of reading (3% of the variance). Grade level accounted for less than 1% of the variance in reading self-concept but for 11% of the variance in the value of reading scale. Previous research with the MRS found similar patterns of scores by sex and grade level. That is, girls have tended to have a more positive reading self-concept and to value reading more than boys while both reading self-concept and value of reading scores decreased as grade level increased (Archambault, Eccles, & Vida, 2010; Applegate & Applegate, 2011; Gambrell et al., 2011; Marinak & Gambrell, 2010; Retelsdorf, Schwartz, & Asbrock, 2014). Reading attitudes have also been found to be more positive for girls than boys and to decrease across grade levels (Kush & Watkins, 1996; McKenna, Conradi, Lawrence, Jang, & Meyer, 2012; McKenna, Kear, & Ellsworth, 1995). Thus, these affective patterns appear to be robust phenomena.

Strong structural validity evidence facilitates both research and practice (Kane, 2013) and should precede research on the relationships between constructs (Meehl, 1990). The current studies have provided strong structural validity evidence for the B-MRS. Thus, research on the relationship of reading self-concept and the value of reading with other theoretical constructs (e.g., reading achievement, reading attitudes, etc.) can now be conducted with these subscales serving as marker variables (Gorsuch, 1988).

Limitations

The biggest limitations of these studies were use of samples of convenience as well as the low response rate of Texas teachers with the resulting inability to judge the representativeness of the normative sample. Although 1,000 Texas teachers were randomly sampled, only 88 responded with B-MRS data from their classrooms. It is possible that the classrooms of these respondents were somehow non-representative. Additionally, lack of demographic information about individual participants makes it impossible to know if the students in the norm sample were representative of the state. Although validity may not be impacted by convenience samples (Mullinix, Druckman, & Freese, 2014), it would be advisable for users to supplement the B-MRS Texas norms with local norms.

Implications for Practice

It is widely accepted that motivation is involved in students' reading development (Afflerbach, Cho, Kim, Crassas, & Doyle, 2013), that reading failure has negative affective correlates (Morgan, Farkas, & Wu, 2012), and that interventions to improve reading

motivation and achievement may be most effective for younger children (Retelsdorf, Köller, & Möller, 2014). Fortunately, investigators (e.g., Baker, Scher, & Mackler, 1997; Edmunds & Bauserman, 2006; Guthrie et al, 2006; Malloy, Marinak, & Gambrell, 2010; Marinak, 2013; McRae & Guthrie, 2009; Meece & Miller, 1999; Monteiro, 2013; Wentzel & Wigfield, 2007; Wigfield, Guthrie, Tonks, & Perencevich, 2004) have identified home and school practices that improve student motivation to read. Several promising school practices were identified by Marinak (2013) who found that fifth grade teachers who offered structured read alouds, cooperative learning via jigsaws, and book club choices improved the perceived value of reading to their students. Other promising school interventions include cross-age peer tutoring, which has improved the reading motivation of both tutors and tutees (Monteiro, 2013) and the concept-oriented reading instruction program (Guthrie, McRae, & Klauda, 2007) that combines reading instruction with support for student motivation. Home practices that have been shown to improve reading motivation include shared storybook reading (Bus, van Ijzendoorn, & Pellegrini, 1995) and other family literacy activities that fostered active parent-child engagement (Yeo, Ong, & Ng, 2014). The B-RMS is a tool that teachers may now systematically apply (e.g., as described by Malloy et al., 2013) to target and monitor interventions that affect reading motivation, knowing that it has exhibited strong evidence of reliability and validity.

References

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print.* Cambridge, MA: MIT Press.

Afflerbach, P. P., & Cho, B.-Y. (2011). The classroom assessment of reading. In M. L. Kamil, P. D. Pearson, E. B. Moje, & P. Afflerbach (Eds.), *Handbook of reading research* (Vol. IV, pp. 487-514). New York, NY: Routledge.

Afflerbach, P., Cho, B.-Y., Kim, J.-Y., Crassas, M. E., & Doyle, B. (2013). Reading: What else matters besides strategies and skills? *The Reading Teacher, 66,* 440-448. doi:10.1002/TRTR.1146

Anderman, E. M., Gray, D. L., & Chang, Y. (2013). Motivation and classroom learning. In W. M. Reynolds & G. E. Miller (Eds.), *Handbook of psychology: Educational psychology* (2nd ed., Vol. 7, pp. 99-116). Hoboken, NJ: Wiley.

Applegate, A. J., & Applegate, M. D. (2011). A study of thoughtful literacy and the motivation to read. *The Reading Teacher, 64,* 226-234. doi:10.1598/RT.64.4.1

Archambault, I., Eccles, J. S., & Vida, M. N. (2010). Ability self-concepts and subjective value in literacy: Joint trajectories from grades 1 through 12. *Journal of Educational Psychology, 102,* 804-816. doi:10.1037/a0021075

Baker, L., Scher, D., & Mackler, K. (1997). Home and family influences on motivations for reading. *Educational Psychologist, 32,* 69-82. doi:10.1207/s15326985ep3202_2

Bear, G. G., Minke, K. M., & Manning, M. A. (2002). Self-concept of students with learning disabilities: A meta-analysis. *School Psychology Review, 31,* 405-427.

Bobko, P., Roth, P. L., & Buster, M. A. (2007). The usefulness of unit weights in creating composite scores. *Organizational Research Methods, 10,* 689-709. doi:10.1177/1094428106/294734

Brown, T. A. (2013). Latent variable measurement models. In T. D. Little (Ed.), *Oxford handbook of quantitative methods: Statistical analysis* (Vol. 2, pp. 257-280). New York, NY: Oxford University Press.

Bus, A. G., van Ijzendoorn, M. H., & Pellegrini, A. D. (1995). Joint book reading makes success in learning to read: A meta-analysis on intergenerational transmission of literacy. *Review of Educational Research, 65,* 1-21. doi:10.3102/00346543065001001

Chamorro-Premuzic, T., Harlaar, N., Greven, C. U., & Plomin, R. (2010). More than just IQ: A longitudinal examination of self-perceived abilities as predictors of academic performance in a large sample of UK twins. *Intelligence, 38,* 385-392. doi:10.1016/j.intell.2010.05.002

Conradi, K., Jang, B. G., & McKenna, M. C. (2014). Motivation terminology in reading research: A conceptual review. *Educational Psychology Review, 26,* 127-164. doi:10.1007/s10648-013-9245-z

De Naeghel, J., & Van Keer, H. (2013). The relation of student and class-level characteristics to primary school students' autonomous reading motivation: A multi-level approach. *Journal of Research in Reading, 36,* 351-370. doi:10.1111/j.1467-9817.2013.12000.x

Eccles, J. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motives* (pp. 75-146). San Francisco, CA: Freeman.

Edmunds, K., & Bauserman (2006). What teachers can learn about reading motivation through conversations with children. *The Reading Teacher, 59,* 414-424. doi:10.1598/RT.59.5.1

Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods, 5,* 155-174. doi:10.1037/1082-989X.5.2.155

Flora, D. B., LaBrish, C., & Chalmers, R. P. (2012). Old and new ideas for data screening and assumption testing for exploratory and confirmatory factor analysis. *Frontiers in Psychology, 3*(55), 1-21. doi:10.3389/fpsyg.2012.00055

Ford, J. K., MacCallum, R. C., & Tait, M. (1986). The application of exploratory factor analysis in applied psychology: A critical review and analysis. *Personnel Psychology, 39,* 291-314. doi:10.1111/j.1744-6570.1986.tb00583.x

Fulmer, S. M., & Frijters, J. C. (2009). A review of self-report and alternative approaches in the measurement of student

motivation. *Educational Psychology Review, 21,* 219-246. doi:10.1007/s10648-009-9107-x

Gambrell, L. B., Hughes, E. M., Calvert, L., Malloy, J. A., & Igo, B. (2011). Authentic reading, writing, and discussion: An exploratory study of a pen pal project. *Elementary School Journal, 112,* 234-258. doi:0013-5984/2011/11202-0002

Gambrell, L. B., Palmer, B. M., Codling, R. M., & Mazzoni, S. A. (1996). Assessing motivation to read. *The Reading Teacher, 49,* 518-533. doi:10.1598/RT.49.7.2

Gorsuch, R. L. (1988). Exploratory factor analysis. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (2nd ed., pp. 231-258). New York, NY: Plenum.

Gorsuch, R. L. (1997). Exploratory factor analysis: Its role in item analysis. *Journal of Personality Assessment, 68,* 532-560. doi:10.1207/s15327752jpa6803_5

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7,* 6-10. doi:0.1177/074193258600700104

Graham, S., & Weiner, B. (1996). Theories and principles of motivation. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 63-84). New York, NY: Macmillan.

Graham, S., & Weiner, B. (2012). Motivation: Past, present, and future. In K. R. Harris, S. Graham, & T. Urdan (Eds.), *APA educational psychology handbook: Theories, constructs, and critical issues* (Vol. 1, pp. 367-397). Washington, DC: American Psychological Association.

Guthrie, J. T. (2011). Best practices in motivating students to read. In L. M. Morrow & L. B. Gambrell (Eds.), *Best practices in literacy instruction* (4th ed., pp. 177-198). New York, NY: Guilford.

Guthrie, J. T., & Coddington, C. S. (2009). Reading motivation. In K. R. Wentzel and A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 503-525). New York, NY; Routledge.

Guthrie, J. T., McRae, A. C., & Klauda, S. L. (2007). Contributions of concept-oriented reading instruction to knowledge about interventions for motivations in reading. *Educational Psychologist, 42,* 237-250. doi:10.1080/00461520701621087

Guthrie, J. T., Wigfield, A., Humenick, N. M., Perencevich, K. C., Taboada, A., & Barbosa, P. (2006). Influences of stimulating tasks on reading motivation and comprehension. *Journal of Educational Research, 99,* 232-245. doi:10.3200/JOER.99.4.232-246

Hairrell, A., Rupley, W., & Simmons, D. (2011). The state of vocabulary research. *Literacy Research and Instruction, 50,* 253-271. doi:10.1080/19388071.2010.514036

Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning.* New York, NY: Routledge.

Hidi, S., & Harackiewicz, J. M. (2000). Motivating the academically unmotivated: A critical issue for the 21st century. *Review of Educational Research, 70,* 151-179. doi:10.3102/00346543070002151

Hox, J. J. (1995). *Applied multilevel analysis.* Amsterdam, The Netherlands: TT-Publikaties.

Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1-55. doi:10.1080/10705519909540118

Hulme, C., & Snowling, M. J. (2011). Children's reading comprehension difficulties: Nature, causes, and treatments. *Current Directions in Psychological Science, 20,* 139-142. doi:10.1177/0963721411408673

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50,* 1-73. doi:10.1111/jedm.12000

Kelley, M. J., & Decker, E. O. (2009). The current state of motivation to read among middle school students. *Reading Psychology, 30,* 466-485. doi:10.1080/02702710902733535

Klauda, S. L., & Wigfield, A. (2012). Relations of perceived parent and friend support for recreational reading with children's reading motivations. *Journal of Literacy Research, 44,* 3-44. doi:10.1177/1086296X11431158

Kush, J. C., & Watkins, M. W. (1996). Long-term stability of children's attitudes toward reading. *Journal of Educational Research, 89,* 315-319. doi:10.1080/00220671.1996.9941333

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6,* 293-323. doi:10.1016/0010-0285(74)90015-2

Lei, P.-W., & Wu, Q. (2012). Estimation in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 164-180). New York, NY: Guilford.

Linnenbrink, E. A., & Pintrich, P. R. (2002). Motivation as an enabler for academic success. *School Psychology Review, 31,* 313-327.

Malloy, J. A., Marinak, B. A., & Gambrell, L. B. (Eds.). (2010). *Essential readings on motivation.* Newark, DE: International Reading Association.

Malloy, J. A., Marinak, B. A., Gambrell, L. B., & Mazzoni, S. A. (2013). Assessing motivation to read: The motivation to read profile-revised. *The Reading Teacher, 67,* 273-282. doi:10.1002/TRTR.1215

Marinak, B. A. (2013). Courageous reading instruction: The effects of an elementary motivation intervention. *Journal of Educational Research, 106,* 39-48. doi:10.1080/00220671.2012.658455

Marinak, B. A., & Gambrell, L. B. (2008). Intrinsic motivation and rewards: What sustains young children's engagement with text? *Literacy Research and Instruction, 47,* 9-26. doi:10.1080/19388070701749546

Marinak, B. A., & Gambrell, L. B. (2010). Reading motivation: Exploring the elementary gender gap. *Literacy Research and Instruction, 49,* 129-141. doi:10.1080/19388070902803795

Marsh, H. W., Xu, M., & Martin, A. J. (2012). Self-concept: A synergy of theory, method, and application. In K. R. Harris, S. Graham, & T. Urdan (Eds.), *APA educational psychology handbook: Theories, constructs, and critical issues* (Vol. 1, pp. 427-458). Washington, DC: American Psychological Association.

Marulis, L. M., & Neuman, S. B. (2013). How vocabulary interventions affect young children at risk: A meta-analytic review. *Journal on Research on Educational Effectiveness, 6,* 223-262. doi:10.1080/19345747.2012.755591

McKenna, M. C., Conradi, K., Lawrence, C., Jang, B. G., & Meyer, J. P. (2012). Reading attitudes of middle school students: Results of a U.S. survey. *Reading Research Quarterly, 47,* 283-306. doi:10.1002/RRQ.021

McKenna, M. C., Kear, D. J., & Ellsworth, R. A. (1995). Children's attitudes toward reading: A national survey. *Reading Research Quarterly, 30,* 934-956. doi:10.2307/748205

McRae, A., & Guthrie, J. T. (2009). Promoting reasons for reading: Teacher practices that impact motivation. In E. H. Hiebert (Ed.), *Reading more, reading better* (pp. 55-76). New York, NY: Guilford.

Meece, J. L., & Miller, S. D. (1999). Changes in elementary school children's achievement goals for reading and writing: Results of a longitudinal and an intervention study. *Scientific Studies of Reading, 3,* 207-229. doi:10.1207/s1532799xssr0303_2

Meehl, P. E. (1990). Why summaries of research on psychological theories are often uninterpretable. *Psychological Reports, 66,* 195-244. doi:10.2466/pr0.1990.66.1.195

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50,* 741-749. doi:10.1037/0003-066X.50.9.741

Möller, J., Pohlmann, B., Köller, O., & Marsh, H. W. (2009). A meta-analytic path analysis of the internal/external frame of reference model of academic achievement and academic self-concept. *Review of Educational Research, 79,* 1129-1167. doi:10.3102/0034654309337522

Monteiro, V. (2013). Promoting reading motivation by reading together. *Reading Psychology, 34,* 301-335. doi:10.1080/02702711.2011.635333

Morgan, P. L., Farkas, G., & Wu, Q. (2012). Do poor readers feel angry, sad, and unpopular? *Scientific Studies of Reading, 16,* 360-381. doi:10.1080110888438.2011.570397

Morgan, P. L., & Fuchs, D. (2007). Is there a bidirectional relationship between children's reading skills and reading motivation? *Exceptional Children, 73,* 165-183. doi:10.1177/001440290707300203

Murphy, P. K., & Alexander, P. A. (2000). A motivated exploration of motivation terminology. *Contemporary Educational Psychology, 25,* 3-53 doi:10.1006/ceps.1999.1019

Mullinix, K., Druckman, J., & Freese, J. (2014). *The generalizability of survey experiments* (Working Paper #14-19). Evanston, IL: Northwestern University Institute for Policy Research.

Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods & Research, 22,* 376-398. doi:10.1177/0049124194022003006

Muthén, B. O. (1997). Latent variable modeling of longitudinal and multilevel data. *Sociological Methodology, 27,* 453-480. doi:10.1111/1467-9531.271034

Muthén, B. O., & Muthén, L. K. (2014). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

National Center for Education Statistics. (2013). *The nation's report card: A first look: 2013 mathematics and reading* (NCES 2014-451). Retrieved from http://nces.ed.gov/nationsreportcard/subject/publications/main2013/pdf/2014451.pdf

Neuhard, R. (2004). *Examining the factor structure of the Motivation to Read Profile among a sample of elementary school students* (Unpublished master's thesis). Pennsylvania State University, State College, PA.

O'Flahavan, J., Gambrell, L. B., Guthrie, J. T., Stahl, S., & Alvermann, D. (1992). Poll results guide activities of research center. *Reading Today, 9*(5), 12.

Paris, S. G., & Oka, E. R. (1986). Self-regulated learning among exceptional children. *Exceptional Children, 53,* 103-108.

Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading, 18,* 22-37. doi:10.1080/10888438.2013.827687

Piasta, S. B., & Wagner, R. K. (2010). Developing early literacy skills: A meta-analysis of alphabet learning instruction. *Reading Research Quarterly, 45,* 8-38. doi:10.1598/RRQ.45.1.2

Pintrich, P. R., & Schunk, D. H. (2002). *Motivation in education: Theories, research, and application* (2nd ed.). Englewood Cliffs, NJ: Merrill/Prentice Hall.

Pitcher, S. M., Albright, L. K., DeLaney, C. J., Walker, N. T., Seunarinesingh, K., Mogge, S., ... & Dunston, P. J. (2007). Assessing adolescents' motivation to read. *Journal of Adolescent & Adult Literacy, 50,* 378-396. doi:l0.1598/JAAl.50.5.5

Quirk, M., Schwanenflugel, P. J., & Webb, M.-Y. (2009). A short-term longitudinal study of the relationship between motivation to read and reading fluency skill in second grade. *Journal of Literacy Research, 41,* 196-227. doi:10.1080/10862960902908467

R Development Core Team. (2014). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Available from http://www.R-project.org.

Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D., & Seidenberg, M. S. (2001). How psychological science informs the teaching of reading. *Psychological Science in the Public Interest, 2,* 31-74. doi:10.1111/1529-1006.00004

Retelsdorf, J., Köller, O., & Möller, J. (2014). Reading achievement and reading self-concept − Testing the reciprocal effects model. *Learning and Instruction, 29,* 21-30. doi:10.1016/j.learninstruc.2013.07.004

Retelsdorf, J., Schwartz, K., & Asbrock, F. (2014, June 16). "Michael can't read!" Teachers' gender stereotype and boys' reading self-concept. *Journal of Educational Psychology.* Advance online publication. http://dx.doi.org/10.1037/a0037107

Runge, T. (1998). *Phonemic awareness skills of preschool children* (Unpublished master's thesis). Pennsylvania State University, State College, PA.

Salvia, J., Ysseldyke, J. E., & Bolt, S. (2010). *Assessment in special and inclusive education* (11th ed.). Belmont, CA: Wadsworth.

Schiefele, U., Schaffner, E., Möller, J., & Wigfield, A. (2012). Dimensions of reading motivation and their relation to reading behavior and competence. *Reading Research Quarterly, 47,* 427-463. doi:10.1002/RRQ.030

Schunk, D. H. (2000). Coming to terms with motivation constructs. *Contemporary Educational Psychology, 25,* 116-

119. doi:10.1006/ceps.1999.1018

Slavin, R. E., Lake, C., Chambers, B., Cheung, A., & Davis, S. (2009). Effective reading programs for the elementary grades: A best-evidence synthesis. *Review of Educational Research, 79,* 1391-1466. doi:10.3102/0034654309341374

Snow, C. E., Burns, M. S., & Griffin, P. (1998). *Preventing reading difficulties in young children.* Washington, DC: National Academy Press.

Snowling, M. J., & Hulme, C. (2011). Evidence-based interventions for reading and language difficulties: Creating a virtuous circle. *British Journal of Educational Psychology, 81,* 1-23. doi:10.1111/j.2044-8279.2010.02014.x

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21,* 360-407. doi:10.1598/RRQ.21.4.1

The Nation's Report Card. (2012). *Reading 2011: National assessment of educational progress at grades 4 and 8* (NCES 2012-457). Washington, DC: U. S. Department of Education.

Uguroglu, M. E. & Walberg, H. J. (1979). Motivation and achievement: A quantitative synthesis. *American Educational Research Journal, 16,* 375-389. doi:10.3102/00028312016004375

Wainer, H. (1976). Estimating coefficients in linear models: It don't make no nevermind. *Psychological Bulletin, 83,* 213-217. doi:10.1037/0033-2909.83.2.213

Watkins, M. W., & Coffey, D. Y. (2004). Reading motivation: Multidimensional and indeterminate. *Journal of Educational Psychology, 96,* 110-118. doi:10.1037/0022-0663.96.1.110

Wentzel, K. R., & Wigfield, A. (2007). Motivational interventions that work: Themes and remaining issues. *Educational Psychologist, 42,* 261-271. doi:10.1080/00461520701621103

Wigfield, A., Guthrie, J. T., & McGough, K. (1996). *A questionnaire measure of children's motivations for reading* (Instructional Resource No. 22). Athens, GA: National Reading Research Center, University of Georgia and University of Maryland.

Wigfield, A., Guthrie, J. T., Tonks, S., & Perencevich, K. C. (2004). Children's motivation for reading: Domain specificity and instructional influences. *Journal of Educational Research, 97,* 299-309. doi:10.3200/JOER.97.6.299-310

Yeo, L. S., Ong, W. W., & Ng, C. M. (2014). The home literacy environment and preschool children's reading skills and interest. *Early Education and Development, 25,* 791-814. doi:10.1080/10409289.2014.862147c

Young, D. (2000). *A construct validity study of the Motivations for Reading Questionnaire* (Unpublished master's thesis). Pennsylvania State University, State College, PA.

Article

Adopting a New Test Edition:
Psychometric and Practical Considerations

A. Alexander Beaujean
Baylor University

When a new edition of a test is published, it is tempting to purchase it as soon as there are funds to do so. In addition to updated norms, the new edition is likely to have administration and scoring advantages over previous editions, and ethics codes seem to mandate the immediate use of newer editions once they are published. Instead, I argue that school psychologists should not immediately adopt a new test edition when it is published, but, instead, should be prudent and conduct a thorough investigation of the new edition before adopting it in their practice. To aid in this process, I outline some basic information that school psychologists should gather in making the test adoption decision. Gathering such information may take considerable time and effort, but this process is aligned with School Psychology best practices and aids school psychologists in providing evidence-based assessment.

Keywords: adopting new test editions, adopting new assessments, school psychology

With the recent release of new editions of the Wechsler Intelligence Scale for Children (Wechsler, 2014) and Woodcock-Johnson (Schrank, McGrew, & Mather, 2014), the issue of test adopting has again become an issue in the practice of School Psychology. Thus, it is a good time for school psychologists to review criteria for choosing if, and when, to adopt a new test edition.

With the publication of a new test edition--especially tests that have wide name recognition and whose previous editions are commonly used--it is tempting to purchase them as soon as there are funds to do so. After all, the advertising for the new editions often touts their scoring and administration advantages, such as improved normative data, better psychometric properties, or more culture fairness. Moreover, ethics codes seem to mandate the use of newer test editions:

> (a) Psychologists do not base their assessment or intervention decisions or recommendations on data or test results that are outdated for the current purpose.
>
> (b) Psychologists do not base such decisions or recommendations on tests and measures that are obso-lete and not useful for the current purpose.

(American Psychological Association, 2010, Standard 9.08)

> If using norm-referenced measures, school psychologists choose instruments with up-to-date normative data. (National Association of School Psychologists, 2010, Standard II.3.2, point 2)

The temptation to adopt new test editions soon after they are released, however, should be resisted by school psychologists. Such behavior does not follow an evidence-based approach to assessment, and may even be unethical (Hunsley & Mash, 2011). In what follows, I describe why school psychologists should not rush to adopt new test versions as well as issues to consider before making the final decision about whether to adopt the new version of a test.

Why Not Immediately Adopt a New Version of a Test?

New Test Editions are Brand New Tests

New editions of a test should be treated as if they are brand new tests, not an updated version of an existing test. Sharing the same developers or

names does not make scores from the two test editions directly comparable. Because there are many reasons why these scores might not be directly comparable (Bracken, 1988; Kaufman, 2010), determining the equivalence of the two editions' scores is an issue that needs to be investigated empirically (Floyd, Clark, & Shadish, 2008).

It would be great if test publishers devoted a section in the technical manuals of revised tests to address the comparability of scores across editions, but this seldom occurs. Instead, scholars not affiliated with the test typically publish this information, sometimes years after the latest edition of the test was released. For example, in a recent study, Benson, Beaujean, and Taub (in press) compared the underlying structure of the second/revised, third, and fourth editions of the Wechsler Adult Intelligence Scale (WAIS). They found that the factor structures of the third and fourth editions were almost identical, but the factor structures of the second/revised and third editions were not. In other words, scores from the second and third editions of the WAIS measure their intended constructs differently, rendering their scores as similar as scores from two completely different tests of cognitive ability. While studies of the same rigor of Benson et al.'s study have not been conducted with other tests, the evidence that has been published shows some similar problems with direct comparison of scores across test editions (e.g., Beaujean & Sheng, 2014).

Score comparability is not only an issue for revisions of cognitive ability tests. In fact, Strauss, Spreen, and Hunter (2000) argued that the problem is even worse for tests of other types of constructs, such as personality. Consequently, it is probably wise for school psychologists to assume that the scores across editions of any test are not directly comparable--unless evidence is reported to indicate otherwise--and should use other methods to compare scores across test editions (Beaujean & McGlaughlin, in press).

There is No Legal or Ethical Mandate to Adopt a New Test Edition Quickly

In spite of the ostensibly strong language in psychology ethics codes, there really is no legal, or even ethical, mandate to adopt a new edition of a test quickly. Most state psychology/education boards typically do not address when to begin using a new edition of a test, or do so in a very vague manner. For example, the Texas State Board of Examiners of Psychologists (2014) rules on test editions (9465.16.b.4-5) just mimic standard 9.08 from the American Psychological Association's ethical guidelines (APA, 2010).

One could interpret APA's ethical standard 9.08 as being explicit in supporting the idea of adopting new test editions quickly. For example, some have argued that new editions of cognitive ability measures should be adopted somewhere between 6-12 months (Dombrowski, 2003) to 2 years (Kranzler & Floyd, 2013) after they are released. Others, however, argue that such a rigid interpretation--and subsequent quick decision-making--is wrong (e.g., Lichtenstein, 2010). For example, Behnke (2005) argued that APA's standard regarding test revisions only directs practitioners to determine the test that is most appropriate for a given purpose. To make this determination, school psychologists need to use their professional knowledge-- gained from their training and experience--in conjunction with their knowledge of how the test scores are going to be used in deciding which test or test version to use.

A casual reading of Standard II.3.2 of the National Association of School Psychologists' ethical guidelines (NASP, 2010) makes it appear that NASP took a firmer stance concerning test revisions than APA. Nonetheless, Behnke's (2005) argument for a non-rigid interpretation likely applies here as well because point 1 of Standard II.3.2 mimics APA's guidelines by stating, "School psychologists select assessment instruments and strategies that are reliable and valid for the child and the purpose of the assessment" (p. 7). Thus, although there has not been any clarification about this point, NASP likely intended a similar interpretation as APA: school psychologists need to determine the test that is most appropriate for a given purpose.

Just because the ethical guidelines afford psychological practitioners some flexibility in selecting instruments, this flexibility should not be abused. For example, using an older edition of a test when a newer edition has been published for 10 years is not aligned with best practice. What the flexibility

grants to practitioners is an adequate amount of time—how-ever long that might be—to determine if they want to adopt a new edition of a test or use an alternative test to assess the construct of interest.

It is Unethical to Adopt a New Test Edition Without Doing the Appropriate Due Diligence

It is unethical to adopt a new version of a test when it is released without first doing the appropriate due diligence to make sure there is sufficient validity and reliability information to interpret the test's scores in the way suggested by the test developer and publisher. The American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (AERA/APA/NCME, 2014) wrote:

> Prior to the adoption and use of a published test, the test user should study and evaluate the materials provided by the test developer. Of particular importance are those that summarize the test's purposes, specify the procedures for test administration, define the intended populations of test takers, and discuss the score interpretations for which validity and reliability/precision data are available. (Standard 9.2)

Just because a test manual states what its test/subtest scores should measure, it does not necessarily mean that is what the scores actually measure. For example, the Arithmetic subtest has been included on various Wechsler intelligence scales for many years. The test authors and publishers have argued that it measures some mix of working memory and fluid reasoning. Nonetheless, multiple studies have found that when the subtest is examined with measures of fluid reasoning, working memory, and math achievement, it is a better measure of quantitative skills than any other ability (Parkin & Beaujean, 2012; Woodcock, 1990).

Another example is bias in the test's scores (i.e., scores work differently across various demographic groups). While test developers often remove individual items that are possibly biased, that does not mean that the test's scores are necessarily unbiased. Determining if bias exists in a test's scores is a complex issue, but one that can be examined empirically (Wicherts, & Dolan, 2010). Oftentimes, these analyses are conducted by independent researchers after the test has been published, making it imperative for school psychologists to remain up-to-date

on the research that supports (or does not support) the use of particular measures for particular purposes.

These previous two examples represent some of the basic information that should be understood before adopting a test. While tests' technical manuals often provide much information about a test's scores, being able to understand all the presented information requires knowledge of psychometrics that is not covered by many training programs-- neither the specialist nor doctoral levels (Perham, 2010; Reynolds, 2011). How can school psychologists then evaluate a test's technical information in order to make appropriate decisions about test adoption?

One option is to rely on the opinions provided by test authors and publishers. This is a good place to start, but it should not be the only source of information used in making the decision. Relying solely on the opinions of test authors/publishers to make a test adoption decision is akin to relying solely on the opinions provided by pharmaceutical companies to make decisions on whether to take their medication. While their information can be valuable, these individuals also have a monetary incentive to sell their product, so they have a conflict of interest.

A second option is to rely on reviews from independent organizations such as the Buros Center for Testing and American Institutes for Research. The Buros Center for Testing publishes the *Mental Measurements Yearbook* (MMY; Nitko, n.d.), which contains reviews of commercially available tests published in English.[1] The reviews are specifically designed to aid test users in the test adoption process. While purchasing an individual MMY volume or subscribing to the electronic version of MMY can be expensive, the *Test Reviews Online* option allows individuals to purchase test reviews of single instruments (marketplace.unl.edu/buros). At the time of writing this article, each test review was $15.

The American Institutes for Research runs the National Center on Response to Intervention (rti4success.org) and National Center on Intensive

---

[1] The Buros Center for Testing publishes a different resource for tests published partly or entirely in Spanish called *Pruebas Publicadas en Español.*

Intervention (intensiveintervention.org). The National Center on Response to Intervention provides reviews of educational screening tests, while the National Center on Response to Intervention provides reviews of academic and behavioral progress monitoring tests. At the time of writing this article, all the test reviews were available freely.

A third option is to rely on peer-reviewed publications about the tests. Some journals (e.g., *Journal of Psychoeducational Assessment, Assessment for Effective Intervention, Journal of Personality Assessment*) regularly publish reviews of tests by individuals who have no conflicts of interest. Depending on when the test review is published, it could contain anything from a review/interpretation of the technical manual to a review of the peer-reviewed literature published on the test. In addition to journals, some books are devoted to reviewing psychological tests for practitioners.

Information about a test from the test publisher is often available as soon as the test is released. Test reviews from journals, books, or independent organizations, however, are only available after the test has been available--sometimes even years after the initial publication. Consequently, if school psychologists do not feel comfortable evaluating the material in a test's technical manual, then they should postpone the test adoption process--even if it is more than a year after the test is published--until they have gathered enough evidence to make an informed decision that they can adequately defend. As noted by Floyd and Norfolk (2014):

> school psychologists should steadfastly evaluate the evidence supporting the use and interpretation of the instruments they employ based on their own reviews and consideration of student characteristics (e.g., age, gender, and race/ethnicity) as well as the intended uses of the instrument. (p. 272)

In the next section, I discuss some of the information that school psychologists should gather before making the decision to adopt a new edition of a test.

What to Examine when Making a Decision on Adopting a New Edition of a Test

The default decision when making a determination about adopting a test--whether it is a first or revised edition--should be *not* to adopt it. Only after gathering sufficient information about the test

should the initial decision be changed. In what follows, I describe some basic information that should be gathered before adopting a new edition of a test in addition to the typical validity and reliability evidence (Wasserman & Bracken, 2003). This is not intended to be an exclusive or ordinal set of priorities, but rather they are only designed to aid in the decision making process. For a useful checklist to aid in making any test adoption decision, see Evers et al. (2013).

Purpose of the Test Revisions

Probably the most important question to ask when evaluating a new test edition is: Why did the test need to be revised? If the purpose was only to update the norming sample, then the revision process should only consist of updating the norms to make them more representative of the population of interest. Such updates typically do not require the publication of a new edition. Instead, they are often done by providing new scoring tables or updating the scoring software in addition to describing the new norming sample. More likely reasons for creating a new edition of a test are: to replace problematic items/subtests that were discovered after the release of the previous edition, to change or update the scoring procedures to align better with new psychometric developments, and/or to change the content or scores to adhere to a new or updated theory of the test's measured constructs.

Replacing problematic items/subtests is a straightforward process that the test developers should address in the technical manual. Changing the scoring procedures is a more complex process as it typically involves moving to a more sophisticated way of scaling the test and creating test scores (e.g., Roid, 2004). With such revisions, test developers/publishers should provide test users a clear description of how to interpret the values from the new scoring procedures as well as how these changes improve the administration and/or scoring of the test (e.g., Jaffe, 2009). In addition, if the new test scores have to be scored by a computer, is there a way to score the test manually as well? If not, then how does the test publisher suggest checking that the software-derived scores are correct?

If the test was revised so that the content or scores better adhere to a new or updated theory of

the test's measured constructs, then test developers need to be very clear in describing the theory and how the revisions allow for the test's scores to be aligned better with the theory (i.e., content validation). In addition, the test developers/publishers need to cite empirical evidence supporting the theory and explain why the revised/new theory is better than the theory used in developing the previous edition.

Adams (2000) noted that a fourth reason for creating a new test edition is economic. That is, the test developer/publisher may believe that the current edition of the test is not competitive enough with other tests designed to measure the same constructs. Thus, a new edition of the test is developed, at least in part, to maintain the test's competitiveness as to improve the test's economic success.

Norming Sample

If a test's scores have norm-referenced interpretations, then the new edition should have an improved normative sample that represents the population of interest better than the previous edition. The test developers should have defined the population of interest and explained how they gathered the norming sample. In gathering the norming sample, the test developers should have paid attention to having appropriate representation of age, sex, ethnicity, education/occupation (of examinees or examinee's parents), and geographic region, which includes both location in the country as well as urban and rural residents. Typically, an entire section of the test's technical manual is devoted to providing this information.

Cost

Not too long ago, major revisions to clinical tests were sufficiently infrequent that acquiring the revised test was typically economically justifiable. Currently, revisions of major clinical tests--especially tests of intelligence--appear at about 10-year intervals. Moreover, the increased costs of newer test materials and supplemental test resources (e.g., scoring software) have increased the amount of money required to spend in adopting the new test edition. Moreover, school psychologists should consider the time costs of administering, scoring,

and interpreting the test scores from the new edition. For those unfamiliar with conducting a cost-benefit/effectiveness analysis of their psychological assessments, see Yates and Taub (2003).

Usefulness of Additional Constructs or Scores

Often, new test editions include a variety of new scores to calculate and interpret. There is often a need for new scores when scores from the current test edition perform unsatisfactory, such as having poor validity evidence, excessive cost, or bias. Just because a construct can be assessed, however, does not mean it should be assessed. When evaluating the new scores and new constructs they measure, school psychologists should consider if the new scores are required or optional as well as how much time they add to the test administration time.

In addition to the time cost, school psychologists should consider whether assessing the new constructs, or interpreting the new scores, are important and psychometrically defensible. The new scores need to have *incremental validity*, which can be thought of as the degree to which the new scores explain or predict the phenomena of interest better than scores from previous editions of the test (Haynes & Lench, 2003). Frequently, additional subscores add very little information beyond that of the test's overall aggregate scores (Beaujean, Parkin, & Parker, 2014; Sinharay, Puhan, & Haberman, 2011) and have limited exchangeability across tests (Floyd, Bergeron, McCormack, Anderson, & Hargrove-Owens, 2005). Nonetheless, the test developers/publishers should discuss the incremental validity evidence of the new scores in the technical manual or other test documentation.

Test Utility

The major purpose for a school psychologist to administer tests is to aid in making a decision (e.g., diagnoses, interventions, vocational placement; Haynes & Lench, 2003). Thus, it is important to know if scores from the new test edition enable accurate decision making about the constructs they measure. Such information can be found by examining, for example, likelihood ratios, sensitivity/specificity values, or receiver operating characteristic curves (Treat & Viken, 2012; Youngstrom,

2012). Ideally, scores on the revised version of the test would lead to more accurate decisions than those from scores from previous editions, although if the accuracy was high with the previous edition's scores it may be sufficient just to duplicate its accuracy.

Clinical Utility

Clinical utility refers to whether the test's scores improve clinical decisions and/or treatment outcomes (Hunsley & Bailey, 1999). It is an amalgam of the test scores' validity, cost-effectiveness, applicability, practicality, transportability across test users and settings, test user and examinee acceptability, and utility (Haynes & Lench, 2003). There is not a specific measure of clinical utility because so much is based on the environment in which a school psychologist practices, but the key aspect is that there is an improvement over current practice. For example, a new test edition may have high clinical utility for a school psychologist working in a newly created post-secondary position with ample time and monetary resources, but few assessments available to use. The same test edition may have low clinical utility for a school psychologist working in a secondary setting who has fewer time and monetary resources, but also has a large array of assessments available.

Summary

Whenever a new test is published, whether it is a first edition or fifth edition, school psychologists should be prudent and conduct an adequate investigation before adopting it for practice. While different editions of a test carry the same name, the editions should be treated as if they are unique tests, with the editions' scores having different properties and not being directly comparable. Moreover, while there is no legal or ethical mandate to adopt a new test edition quickly, there is an ethical mandate to conduct the appropriate due diligence before adopting a test for use in clinical practice.

To aid in making the decision to adopt a new test edition, I outlined some basic information that school psychologists should gather: the purpose of the revisions, norming sample, cost, usefulness of any additional constructs or scores, test utility, and

clinical utility. While gathering such information may take considerable time and effort, it is better than the alternative of adopting a new test edition too quickly then finding out later that there are major problems with some of the test's scores, or having to explain in court the lack of due diligence in selecting the test (Reynolds & Milam, 2011). Moreover, gathering the appropriate evidence before deciding to adopt a new test edition is not only aligned with best practice, but it allows for a school psychologist to provide the best assessments to help children and youth succeed academically, socially, behaviorally, and emotionally.

References

Adams, K. M. (2000). Practical and ethical issues pertaining to test revisions. *Psychological Assessment, 12*, 281-286. doi:10.1037/1040-3590.12.3.281

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing* (4th ed.). Washington, DC: Authors.

American Psychological Association. (2010). *Ethical principles of psychologists and code of conduct*. Washington, DC: Author.

Beaujean, A. A., & McGlaughlin, S. M. (in press). Comparing test scores using information from external validity studies. *Applied Neuropsychology: Child.* doi:10.1080/21622965.2014.976863

Beaujean, A. A., Parkin, J., & Parker, S. (2014). Comparing Cattell-Horn-Carroll models for predicting language achievement: Differences between bi-factor and higher-order factor models. *Psychological Assessment, 26,* 789-805. doi:10.1037/a0036745

Beaujean, A. A., & Sheng, Y. (2014). Assessing the Flynn effect in the Wechsler scales. *Journal of Individual Differences, 35*, 63-78. doi:10.1027/1614-0001/a000128

Behnke, S. (2005). Thinking ethically as psychologists: Reflections on ethical standards 2.01, 3.07, 9.08 and 10.04. *Monitor on Psychology, 36*(6). Retrieved from http://www.apa.org/monitor/jun05/ethics.aspx

Benson, N., Beaujean, A. A., & Taub, G. E. (in press). Using score equating and measurement invariance to examine the Flynn effect in the Wechsler Adult Intelligence Scale. *Multivariate Behavioral Research.*

Bracken, B. A. (1988). Ten psychometric reasons why similar tests produce dissimilar results. *Journal of School Psychology, 26,* 155-166. doi:10.1016/0022-4405(88)90017-9

Dombrowski, S. C. (2003). Ethical standards and best practices in using newly revised intelligence tests. *NASP Communiqué, 32*(1), 12-13.

Evers, A., Hagemeister, C., Høstmælingen, A., Lindley, P. A., Muñiz, J. e., & Sjöberg, A. (2013). *EFPA review model for the description and evaluation of psychological and educa-*

tional tests (Version 4.2.6). Brussels, Belgium: European Federation of Psychology Associations. Retrieved from http://www.efpa.eu/professional-development

Floyd, R. G., Bergeron, R., McCormack, A. C., Anderson, J. L., & Hargrove-Owens, G. L. (2005). Are Cattell-Horn-Carroll (CHC) broad ability composite scores exchangeable across batteries? *School Psychology Review, 34,* 329-357.

Floyd, R. G., Clark, M. H., & Shadish, W. R. (2008). The exchangeability of IQs: Implications for professional psychology. *Professional Psychology: Research and Practice, 39,* 414-423. doi:10.1037/0735-7028.39.4.414

Floyd, R. G., & Norfolk, P. A. (2014). Best practices in identifying, evaluating, and communicating research evidence. In P. L. Harrison & A. Thomas (Eds.), *Best practices in school psychology: Foundations* (pp. 265-280). Bethesda, MD: National Association of School Psychologists.

Haynes, S. N., & Lench, H. C. (2003). Incremental validity of new clinical assessment measures. *Psychological Assessment, 15,* 456-466. doi:10.1037/1040-3590.15.4.456

Hunsley, J., & Bailey, J. M. (1999). The clinical utility of the Rorschach: Unfulfilled promises and an uncertain future. *Psychological Assessment, 11,* 266-277. doi:10.1037/1040-3590.11.3.266

Hunsley, J., & Mash, E. J. (2011). Evidence-based assessment. In D. H. Barlow (Ed.), *The Oxford handbook of clinical psychology* (pp. 76-97). New York, NY: Oxford University Press.

Jaffe, L. E. (2009). *Development, interpretation, and application of the W score and the relative proficiency index* (Woodcock-Johnson III Assessment Service Bulletin 11). Rolling Meadows, IL: Riverside Publishing.

Kaufman, A. S. (2010). "In what way are apples and oranges alike?" A critique of Flynn's interpretation of the Flynn effect. *Journal of Psychoeducational Assessment, 28,* 382-398. doi:10.1177/0734282910373346

Kranzler, J. H., & Floyd, R. G. (2013). *Assessing intelligence in children and adolescents: A practical guide.* New York, NY: Guilford.

Lichtenstein, R. (2010). How soon must you switch to the new version of a test? *NASP Communiqué, 38*(8), 14-15.

National Association of School Psychologists. (2010). *Principles for professional ethics.* Washington, DC: Author.

Nitko, A. J. (n.d.). Using a Mental Measurements Yearbook Review and other materials to evaluate a test. Retrieved from http://buros.org/using-mental-measurements-yearbook-review-and-other-materials-evaluate-test

Parkin, J., & Beaujean, A. A. (2012). The effects of Wechsler Intelligence Scale for Children-Fourth Edition cognitive abilities on math achievement. *Journal of School Psychology, 50,* 113-128. doi:10.1016/j.jsp.2011.08.003

Perham, H. (2010). *Quantitative training of doctoral school psychologists: Statistics, measurement, and methodology curriculum.* (Unpublished master's thesis). Arizona State University, Tempe, AZ.

Reynolds, C. R. (2011). Perspectives on specialization in school psychology training and practice. *Psychology in the Schools, 48,* 922-930. doi:10.1002/pits.20598

Reynolds, C. R., & Milam, D. A. (2011). Challenging intellectual testing results. In D. Faust (Ed.), *Coping with psychiatric and psychological testimony* (6th ed., pp. 311-334). New York, NY: Oxford University Press.

Roid, G. H. (2004). *Quality of performance and change-sensitive assessment of cognitive ability.* Paper presented at the annual meeting of the International Test Users' Conference, Melbourne, Australia. Retrieved from http://research.acer.edu.au/research_conferenceITU_2004/4

Schrank, F. A., McGrew, K. S., & Mather, N. (2014). *Woodcock-Johnson IV.* Rolling Meadows, IL: Riverside.

Sinharay, S., Puhan, G., & Haberman, S. J. (2011). An NCME instructional module on subscores. *Educational Measurement: Issues and Practice, 30*(3), 29-40. doi:10.1111/j.1745-3992.2011.00208.x

Strauss, E., Spreen, O., & Hunter, M. (2000). Implications of test revisions for research. *Psychological Assessment, 12,* 237-244. doi:10.1037/1040-3590.12.3.237

Texas State Board of Examiners of Psychologists. (2014). *Texas State Board of Examiners of Psychologists acts and rules.* Austin, TX: Author.

Treat, T. A., & Viken, R. J. (2012). Measuring test performance with signal detection theory techniques. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf & K. J. Sher (Eds.), *APA handbook of research methods in psychology, Vol 1: Foundations, planning, measures, and psychometrics* (pp. 723-744). Washington, DC: American Psychological Association.

Wasserman, J. D., & Bracken, B. A. (2003). Psychometric characteristics of assessment procedures. In J. R. Graham & J. A. Naglieri (Eds.), *Handbook of psychology (Vol 10): Assessment psychology* (pp. 43-66). Hoboken, NJ: John Wiley and Sons.

Wechsler, D. (2014). *Wechsler intelligence scale for children* (5th ed.). San Antonio, TX: The Psychological Corporation.

Wicherts, J. M., & Dolan, C. V. (2010). Measurement invariance in confirmatory factor analysis: An illustration using IQ test performance of minorities. *Educational Measurement: Issues and Practice, 29*(3), 39-47. doi:10.1111/j.1745-3992.2010.00182.x

Woodcock, R. W. (1990). Theoretical foundations of the WJ-R measures of cognitive ability. *Journal of Psychoeducational Assessment, 8,* 231-258. doi:10.1177/073428299000800303

Yates, B. T., & Taub, J. (2003). Assessing the costs, benefits, cost-effectiveness, and cost-benefit of psychological assessment: We should, we can, and here's how. *Psychological Assessment, 15,* 478-495. doi:10.1037/1040-3590.15.4.478

Youngstrom, E. A. (2012). Future directions in psychological assessment: Combining evidence-based medicine innovations with psychology's historical strengths to enhance utility. *Journal of Clinical Child and Adolescent Psychology, 42,* 139-159. doi:10.1080/15374416.2012.736358

Article

The Comparative Effectiveness of Two Addition Math Fact Interventions: Modified Incremental Rehearsal and Cover, Copy, Compare

Kathryn E. Jaspers
Portland Public Schools

Noga Sundberg, Gail M. Cheramie, Jaymie Hoffman
University of Houston – Clear Lake

Erin E. C. Henze
University of Detroit Mercy

This study compared two math fact interventions in a 9-year-old student with an intellectual disability. The first intervention, Cover, Copy, Compare (CCC), required the student to look at a problem and answer, cover it, copy it, and compare it to the original stimulus. The second intervention, a modified Incremental Rehearsal (mIR) procedure, used flashcards to practice addition facts by first teaching the student an unknown fact and then incrementally practicing the unknown fact among known facts. Results indicated that the student performed better on flashcard assessments in mIR than the CCC or control conditions, but results did not initially generalize to paper/pencil assessments. After a second intervention phase with additional generalization practice, the student also performed better in the mIR than the CCC or control conditions on the paper/pencil assessments. Results were maintained at 4-week follow-up.

Keywords: math, fluency, accuracy, Cover Copy Compare, Incremental Rehearsal

Although national scores in mathematics have improved in the past decade, the 2011 National Assessment of Educational Progress (NAEP) findings showed that 60% of fourth-grade students and 65% of eighth-grade students did not meet the required proficiency level of mathematic skills (National Center for Education Statistics, 2011). Additionally, the National Mathematics Advisory Panel indicated in its report from 2008 that American students cannot solve basic math facts problems as efficiently as students from other countries (NMAP, 2008). Students whose math skills are below the appropriate grade level are at risk for negative outcomes, including failure to meet state-derived benchmarks, retention and dropout (Rhymer, Dittmer, Skinner, & Jackson, 2000). The implications of these findings highlight the need for mathematics intervention research to provide school professionals with empirically supported interventions in order to effectively address students' mathematics difficulties.

Mathematics proficiency is an important aspect of everyday life. We encounter the need for basic math operations when using money or organizing our time. Many educational opportunities and jobs require a high level of mathematical know-

ledge. Mathematical proficiency is made up of five parts: conceptual understanding, procedural fluency, strategic competence, adaptive reasoning, and productive disposition (Kilpatrick, Swafford, & Findell, 2001). According to Burns et al. (2012), although all five areas of mathematical proficiency are important, of particular importance is fluent computation (i.e., procedural fluency), indicated as a goal by both the National Council of Teachers of Mathematics (NCTM) and NMAP. Fluent computation, or the ability to perform basic math operations accurately and quickly, is a building block in the development of math proficiency.

Fluent computation consists of both accurate and rapid responding. The means by which fluent computation is achieved can be best understood using Haring and Eaton's (1978) instructional hierarchy. According to the instructional hierarchy, students develop skills in a progression of learning stages. The first stage of the hierarchy is the acquisition stage, which focuses on responding accurately to a task. The second stage in the instructional hierarchy is fluency building. Once the skill is acquired it is necessary for one to become fluent so that the skill will become functional in the individual's environment. Fluency can be developed by methods of drill, practice, and use of reinforcement. The third stage is generalization. The goal of this stage is to apply a previously learned skill to a new set of stimuli that is similar to those used in instruction. For example, a student may generalize a math problem learned vertically on a worksheet when he or she is presented with a vertical presentation of the problem or when presented with the problem on a computer program. The final stage of the hierarchy is application or adaption. The goal of this stage is for the student to be able to adapt, modify, or apply a previously learned response to new and different stimuli. For example, it would be adaption if a student is taught 5x7, 5x8, and 5x9 and he or she uses this knowledge to answer 5x10 and 5x11 (Haring & Eaton, 1978).

Based on the instructional hierarchy, students first need to become fluent in their computation of math operations before developing an understanding of more complex and abstract concepts that are necessary for math reasoning and problem solving (Axtell, McCallum, Bell, & Poncy,

2009; Codding, Hilt-Panahon, Panahon, & Benson, 2009). Information processing theory supports the view that without the ability to retrieve facts directly or automatically, students are likely to experience a high cognitive load which could impede their learning of more complex tasks (Cumming & Elkins, 1999). Furthermore, students who respond more fluently generally maintain those math skills over time, are more successful in applying those skills to new mathematic problems (Rhymer et al., 2000), and exhibit lower levels of anxiety about math (Cates & Rhymer, 2003).

Previous Research

As noted above, basic skills and proficiency in mathematics are an important aspect of everyday life, and this is especially true for students with intellectual disability (ID). Butler, Miller, Lee, and Pierce (2001) conducted a literature review of math interventions for students with mild to moderate ID. The review spanned a 10-year period from 1989-1998 and 16 articles that collectively included 271 students were analyzed. Single-subject designs were used in the majority of the studies (12 of 16) and 10 studies involved computation skills. Butler et al. noted that the emphasis in mathematics instruction had shifted from basic skills (e.g., counting, numeral recognition) to computation and problem-solving. The authors concluded that students with ID benefited from interventions that involved "frequent feedback, explicit instruction and ample drill-and-practice" (p. 29).

Browder, Spooner, Ahlgrim-Delzell, Harris, and Wakeman (2008) conducted a review of 68 studies published between 1975 and 2005 that collectively included 493 individuals. The majority of the studies reviewed applied a single-subject design (54 of 68 studies) and the majority of students within the studies (336 of 493) displayed moderate ID. The evidence-based practice of systematic instruction (i.e., clearly defined goals, prompting, feedback, prompt-fading) was associated with the best outcomes.

The results of the reviews indicated that students with mild to moderate levels of ID can learn specific computation skills. The elements of evidence-based interventions involve intensive, individualized, systematic instruction coupled with

repeated practice and frequent feedback. The focus of this investigation was on specific individualized interventions to teach math facts. The interventions are intensive, delivered to individuals in a one-to-one instructional arrangement, involve repeated practice and provide for frequent feedback.

Several interventions that improve both accuracy and fluency of math facts have been developed (Burns, 2005; McCallum, Skinner, & Hutchins, 2004; Skinner, Turco, Beatty, & Rasavage, 1989). Many of these interventions use immediate feedback to promote accurate responses and prevent students from practicing errors (Skinner et al., 1989). To increase automaticity, these interventions allow for many practice opportunities in a brief period of time (Skinner, McLaughlin, & Logan, 1997). Two such interventions are Cover, Copy, and Compare (CCC) and Incremental Rehearsal (IR).

Cover, Copy, and Compare

Cover, Copy, and Compare (CCC), an intervention initially designed to improve spelling, has been shown to be effective across curricula, settings and subjects. It was adapted by Skinner et al. (1989) for improving accurate and fluent responding across a variety of mathematics calculation skills. The basic CCC method is comprised of the following steps: (a) the student looks at an academic stimulus (e.g., a math problem and its answer) and is instructed to study it, (b) the student covers the problem and answer on the left side of the page, and then writes the problem and answer on the right side of the page, and (c) the student evaluates the response by comparing it to the stimulus item (Skinner et al., 1997). If the problem and answer were written correctly, the student proceeds to the next item and repeats the CCC procedure. If the response was incorrect, the student engages in an error correction procedure, such as re-writing the correct response one or more times. Adaptations of CCC have been shown to be effective in increasing math accuracy and fluency. Several studies have altered the form of the student response by asking the student to verbalize the response (either out loud or sub-vocally) instead of writing the response (Skinner, Bamberg, Smith, & Powell, 1993; Skinner, Ford & Yunker, 1991).

Studies have also demonstrated the efficacy of CCC in a classroom setting (Poncy, Skinner, & O'Mara, 2006), as well as at home (Stading, Williams, & McLaughlin, 1996). Other studies have incorporated performance feedback and goal setting in order to increase the effectiveness of the intervention (Codding, Eckert, Fanning, Shiyko, & Solomon, 2007; Skinner et al., 1993). Rapid CCC learning trial rates also allow time for students to engage in over-learning which has been shown to lead to better maintenance (Skinner et al., 1997).

Incremental Rehearsal

Incremental Rehearsal (IR) is a technique that was first developed in order to improve vocabulary skills (Tucker, 1988). IR is a flashcard drill method in which known items are interspersed among unknown items at a ratio of one unknown to five to nine known items (Burns & Boice, 2009; Burns, Zaslofsky, Kanive, & Parker, 2012). The known items are interspersed among the unknown items with increasingly more known items being presented, until the unknown item has been presented five to nine times depending on student factors such as age and attention span (e.g., *unknown item*, first known item; *unknown item*, first known item, second known item; *unknown item*, first known item, second known item, third known item, and so on; Burns & Sterling-Turner, 2010). After this sequence, if the student correctly answers/identifies the item, the unknown item is considered known, one of the previous known items is removed, a new unknown item is introduced, and the process is repeated (Burns & Boice, 2009).

Previous studies have found IR to be effective for teaching words, letter sounds, and vocabulary (Burns et al., 2012). Although IR has produced promising results, there are only a few studies that show the effectiveness of IR for teaching math facts, as most of the research addresses acquisition and retention of reading words and includes children without disabilities (Burns, 2005). A recent meta-analysis of IR research in various intervention areas found only two studies that researched the effectiveness of IR with math facts (Burns et al., 2012). One of these studies showed IR was effective in teaching multiplication facts to children with a learning disability in

mathematics (Burns, 2005) and the other was an extension of the Burns (2005) study which supports the findings of the previous study and provides evidence that IR can improve generalization of skills to similar stimulus conditions (Codding, Archer, & Connell, 2010).

Students whose academic skills are delayed are typically not learning at an acceptable rate; thus, instructional time is an important variable when evaluating the relative effects of interventions (Skinner, Belfiore, & Watson, 2002). Interventions that are more efficient (take less time) allow for more learning trials in a set amount of time. A higher number of learning trials has been shown to increase learning levels during the acquisition (Albers & Greer, 1991), fluency building (Skinner & Shapiro, 1989), and maintenance (Ivarie, 1986) stages of learning (Skinner, Belfiore, Mace, Williams-Wilson, & Johns, 1997). When searching for an effective intervention to remediate academic skills problems, educators are interested in knowing which intervention provides the most learning in the least amount of time (i.e., learning rate). However, many treatment comparison studies fail to take into account the amount of time spent on each intervention (Poncy, Skinner, & McCallum, 2012).

Purpose

CCC has been shown to improve math fact fluency with students in general education and students with mild learning disabilities; one study researched the effectiveness of CCC on math fact accuracy and fluency in a student functioning at a cognitive level below that of mild mental retardation (Poncy, Skinner, & Jaspers, 2007). IR has been shown to be effective for teaching words, letter sounds and vocabulary, but only two studies have been conducted to show its effectiveness with math facts (Burns et al., 2012). The current study extends the research by comparing the effects of these two interventions (CCC and IR) on basic math fact accuracy and fluency in a student with an intellectual disability. Furthermore, this study adds to the research by implementing a modified procedure of IR, when the student begins with few or no known problems.

Method

Participant

The participant in this study was Lilly (pseudonym), a 9-year-old Caucasian female who had recently completed 2nd grade. She was identified through her public school as a student with an Intellectual Disability (ID) with a secondary classification of Speech Impairment (SI). Lilly's most recent evaluation was completed toward the end of 1st grade. Lilly's Full Scale IQ score at the time of evaluation was 55. Lilly was assessed in math as part of a comprehensive evaluation conducted by the school district. On the Kaufman Test of Educational Achievement – Second Edition (KTEA-II), Lilly obtained a Math Computation standard score of 68. Related to the present study, Lilly's Individualized Education Plan (IEP) objectives in math included showing quantities of 0-10 using sets of concrete objects and adding and subtracting 1-digit number sentences with manipulatives and visual cues.

Upon approval from the university's human subjects review board, a flyer was given to three directors of special education within close geographical proximity to a university clinic, and each director disseminated the flyer to the special education department chair at several elementary schools within their district. The flyer requested students with mild intellectual disability or low cognitive functioning who had recently finished second through fifth grades for summer math intervention research. The flyers were then given to parents by either the department chair or teacher at the elementary school and parents contacted the researchers. Parents provided a copy of the student's most recent evaluation, and parent consent and child assent were received. The parents were given $100 and the student was given a $15 gift certificate upon completion of the study.

The study was conducted at a university clinic that has one-way observation capacity. Two graduate student researchers served as interventionists. Both graduate students had completed their first year of a three-year specialist program in school psychology. The first author trained both graduate students in the implementation of the intervention and measurement procedures, and the

graduate students demonstrated correct implementation (100% accuracy) in practice sessions prior to beginning the study.

Materials

Materials for this study included flashcards, curriculum-based measures (CBMs), CCC worksheets, and generalization worksheets. Additional materials included a stopwatch, pencils, an index card (to cover the stimulus during CCC), and a half sheet of colored cardstock paper (for the CCC generalization, described later in the procedures section).The flashcards were 3 inches by 4 inches. Each flashcard had a vertical addition problem on one side with the completed problem on the back.

On the experimenter-created CBMs, facts were presented vertically on an 8 x 11 in. sheet of paper, with each fact repeated twice. The experimenters constructed six versions of the assessments for each of the three conditions (i.e., CCC, mIR, control). For all assessments, no one problem was repeated in the same line or directly above/underneath the same problem. Scripted directions were used (Poncy et al., 2007), and Lilly was given 1 min to work on each CBM.

Each CCC worksheet contained a grid of 30 boxes (five columns by six rows) presented on an 8 x 11 in. sheet of paper. The first column contained the target fact. The second column contained a plus sign and an equal field (to reduce writing demands) and a small circle in the corner for Lilly to check if she copied the problem correctly. The third, fourth, and fifth columns were empty for repeated practice.

The generalization worksheets were similar to the CBMs (8 x 11 in. sheet with vertical addition problems) but with only the six targeted math facts that had been practiced during that session, repeated three times. No two math facts were immediately repeated, on the same row, or directly above/ underneath the same problem. Six versions of the generalization worksheet were developed for each of the two experimental conditions: Cover, Copy, Compare (CCC) and modified Incremental Rehearsal (mIR).

Experimental Design and Variables

An adapted alternating treatments design (Sindelar, Rosenberg, & Wilson, 1985) was used to evaluate Lilly's math fact fluency and automaticity under the three conditions: mIR, CCC, and control. This design was selected in order to control for testing effects, history effects (specifically, learning outside of experimental conditions), and spillover effects between conditions (Sindelar et al., 1985; Skinner & Shapiro, 1989). Time was held constant across interventions to provide a comparison of learning rates (Skinner, 2008), with each intervention lasting 8 min. Because time was held constant, trials per session (opportunities to respond) varied. In an alternating treatment design, conditions are counterbalanced to control for sequencing effects.

Dependent variables. Because previous research has shown that math fact learning does not always transfer to different response topographies (e.g., Duhon, House, & Stinnett, 2012), both written and flashcard measures were used as dependent variables. There were three dependent variables assessed in this study. From the 1-min CBMs, digits correct per minute (DCM) and percentage of attempted problems correct per minute (%CM) were calculated in order to assess both fluency and accuracy. For the DCM variable, digits were counted as correct when the correct digit was written in the correct column (Shapiro, 2011). For example, the problem $8+2 = 10$ includes two correct digits, an answer of 11 includes one correct digit (the tens column), an answer of 20 includes one correct digit (ones column), and an answer of 9 includes zero correct digits. Also from the CBMs, %CM was calculated by taking the number of problems correct and dividing by the number of problems attempted. Problems that were not attempted were not counted as correct or incorrect.

Flashcard automaticity was the third dependent variable. At the beginning of each session, the experimenter presented each problem on a flashcard for 2 s by laying the flashcard flat on the table in front of the participant. Any problem correctly answered within 2 s was counted as correct. Incorrect answers or correct answers beyond 2 s were counted as incorrect.

Independent variables. Three conditions were included in the study: Cover, Copy, Compare (CCC), modified Incremental Rehearsal (mIR) and control. Each condition was randomly assigned to a set of addition math facts. The sets were developed by first dividing addition facts (sums to 10) into three mutually exclusive sets using stratified random assignment. Two facts were omitted (6+4 and 0+0) to create three equivalent (i.e., of equal difficulty) sets of 11 facts each (see Table 1 for sets). The control condition was not targeted through any intervention but was assessed each day during all phases to measure testing, spillover, and history effects (Sindelar et al., 1985; Skinner & Shapiro, 1989).

Because the experimenters deemed the total set sizes of 11 problems as too large for Lilly to practice each session, set sizes of six unmastered problems were used for each intervention in each session. It should be noted that the terms 'known' and 'mastered' are not used interchangeably in this study. Consistent with previous IR research, the terms 'known' and 'unknown' are used to describe words *within* a session, and items can go from unknown to known within a session by answering the item correctly (see Burns et al., 2012). However, it may be possible that the item that became 'known' within a session using IR criteria was not truly mastered (i.e., recalled at a later time). See examples of each term below. For the CCC condition, Lilly practiced only unmastered problems. Problems were considered mastered when they were answered correctly on two consecutive sessions of flashcard assessments. Once a problem was mastered, it was removed from the practice set and the next problem from the larger set of 11 was added, keeping the session set size at six (for a description of flow lists, see McLaughlin, Reiter, Mabee, & Byram, 1991). In the mIR condition, unknown problems were practiced one at a time while interspersed with up to five known problems, for a total set size of six. When an unknown problem was learned within a session (i.e., practiced correctly upon its final administration in the mIR sequence), it became a known problem, and a known problem was removed from the practice set. All assessments were conducted with all 11 problems from the set, regardless of whether each problem was practiced that session.

CCC	IR	Control
0+1	2+0	0+3
6+0	0+5	4+0
0+7	9+0	0+8
1+3	1+1	1+2
6+1	1+7	4+1
8+1	5+1	9+1
2+2	5+5	3+3
7+2	7+3	8+2
3+4	3+2	3+5
6+3	2+6	2+5
5+4	4+2	4+4

Table 1. Math Facts

Cover, copy, compare (CCC). For the CCC intervention, Lilly first stated the problem and answer aloud. She then covered the stimulus problem and answer with an index card, wrote the problem and answer in the next column, and then uncovered the stimulus and compared her written response to the stimulus. If correctly written, Lilly placed a check mark next to the problem and stated the problem and answer again. If incorrectly written, she wrote the problem and answer (with original uncovered) three times as an overcorrection technique, verbalizing the problem and answer after each written response. During the CCC condition, an experimenter sat next to Lilly to ensure that she followed the steps correctly and provided prompts to follow the steps, as needed.

Modified incremental rehearsal (mIR). The mIR intervention involved incrementally practicing unknown items among known items. Knowns and unknowns were determined each session through the flashcard assessment at the beginning of each session, which is described in the next section. Typically during incremental rehearsal procedures, known and unknown facts are interspersed in the following manner: *unknown item*, first known item; *unknown item*, first known item, second known item; *unknown item*, first known item, second known item, third known item, and so on until the unknown item is interspersed with five to nine known items (the number of items in the full set can depend on child's age, ability, and specifics

of target set). Once the item has been interspersed with five to nine known items it is considered known, a prior known is removed from the known stack and replaced with the new known, and a new unknown problem is targeted (Burns, 2005; Burns & Sterling-Turner, 2010).

Due to Lilly's age, cognitive ability, and academic history, the experimenters used set sizes of six items (5 known and 1 unknown). However, at the beginning of each session Lilly often had less than five known problems in the set; thus, the incremental rehearsal procedure was modified to provide an equivalent number of opportunities to respond (OTRs) as Lilly would have experienced if there had been five known items in the set. For example, if Lilly knew two problems in the item set, the pattern would look like this (U=unknown, K=known): U1 demonstration and practice; U1; K1; U1, K1, K2; U1, K1, K2; U1, K1, K2; U1, K1, K2. At that point, U1 became K3, and a new unknown was introduced. Regardless of the number of items in her practice set, she was provided with 6 OTRs for each unknown item before it became known. For all items in the mIR condition, Lilly was given 2 s to verbally give a correct response. If she did not respond or responded incorrectly, the experimenter stated the problem and correct answer and had Lilly answer the item again (e.g., "one plus one equals two, what does one plus one equal?").

For the first session, Lilly did not correctly answer any problems during the daily flashcard assessment; therefore there were no known problems when the session began. In this case, U1 was presented six times consecutively with a 2 s pause in between each presentation, at which point it was considered known. The previously described procedures were then followed, with U1 becoming K1. For all other intervention sessions, Lilly correctly answered at least two problems during the flashcard assessment.

Procedure

Sessions occurred three times a week for six weeks. Baseline data were gathered in the first three sessions. After the third session, interventions were modeled and practiced with Lilly using subtraction problems (to limit exposure to targeted problems while providing procedural practice).

Lilly began the interventions in the fourth session, and participated in a total of 15 intervention sessions. A follow-up assessment was conducted 4 weeks after the last intervention session to measure maintenance of skills over time. Across all phases, conditions were counterbalanced to control for sequencing effects. Each phase is described in more detail below.

Baseline phase. At the start of each session, flashcards were used to measure automatic knowledge of all math facts in the three sets. The math facts within each set were randomized by shuffling between each session. The experimenter-constructed addition CBMs from the three conditions were then administered in counter-balanced order. Because the flashcards during the intervention phase (described below) were administered prior to the interventions and the CBMs were administered after the interventions, there were four days of baseline data for the flashcards but only three days for CBM data (DCM and %CM).

Intervention phase. Each session during the intervention phase consisted of three activities in the following order: flashcard assessment, two interventions, and the CBMs. At the beginning of each intervention session, flashcards for all facts from each set were used to assess accurate, automatic recall of the facts. A fact was considered automatic if Lilly gave a correct response within 2 s of its presentation. Flashcards were presented at the beginning of the session because the participant's performance on the flashcard assessment was used to determine known and unknown problems for that session's interventions. Following the flashcards, the interventions were administered, with a short break in between. Following the two interventions, the three CBMs (one from each intervention and a control) were administered.

Intervention + generalization phase. After six intervention sessions, the experimenters noticed that Lilly was becoming more accurate in her practice during each intervention, but she was not transferring those improvements to assessments that required a different topography (i.e., mIR to paper/pencil assessment; CCC to flashcards or paper/pencil assessments). At this point, the

experimenters adjusted the intervention to add a second intervention phase, intervention + generalization. Each session during the intervention + generalization phase began the same way as the previous phase, but after each intervention was finished, Lilly spent 2 min engaging in extra generalization practice with the six unknown math facts from each set (e.g., 8-min CCC intervention, 2-min CCC generalization practice, 8-min mIR intervention, 2-min mIR generalization practice) prior to the session's CBMs.

For the generalization practice, Lilly was given a generalization worksheet, which was similar in appearance to the CBMs, but included only the six problems targeted that session. Lilly was given 2 min to practice as many problems as she could while receiving corrective feedback from the experimenter. If Lilly responded incorrectly, the researcher referred her to the intervention stimuli. The procedure for correcting responses during the generalization practice differed depending on the condition.

Following the CCC intervention, the CCC worksheet was next to the participant as she worked on the generalization worksheet. However, the current problem that was being attempted was covered with card stock so that the answer could not be seen. If Lilly answered incorrectly, she was instructed to look at the CCC problem and try to remember the answer. If she did not remember, the answer was revealed and she recorded the correct answer before attempting the next problem.

Following the mIR intervention, the flashcards were laid in front of the participant, above the generalization worksheet. When the participant responded incorrectly, she was instructed to look at the corresponding flashcard fact and try to remember the problem. If, after that, Lilly did not remember the problem, the flashcard was turned over to reveal the answer. Then the participant recorded the correct answer on the generalization worksheet before attempting the next problem.

Treatment Integrity, Interscorer Agreement, and Treatment Acceptability

Treatment integrity data were gathered by the first author for 37% (7 of 19) of sessions across all phases through a one-way mirror. For each assessment and intervention activity, an implementation checklist was used to measure if all steps were followed and correct feedback given (when applicable). Treatment integrity for CBMs averaged 99% across sessions (range 94-100%), for flashcard assessments was 100% in all sessions, for mIR intervention averaged 98% (range 94-100%), for CCC intervention was 100% in all sessions, for CCC generalization averaged 97% (range 94-100%), and for mIR generalization was 100%. For interscorer agreement, 25% of all assessment probes were scored by a second rater. Interscorer agreement averaged 99% (range 94-100%). Upon completion of the study, Lilly was asked which intervention she liked best (she preferred CCC) and which intervention helped her learn more math problems (she endorsed mIR).

Results

Figure 1 displays data for Lilly on flashcards correct, Digits Correct per Minute (DCM) on assessment sheets, and percentage of attempted problems correct per minute (%CM) on assessment sheets.

Flashcard Problems Correct

Visual analysis of baseline data indicates stable and low performance across all three sets, indicating that most problems in all conditions were unknown prior to the intervention phase. Upon implementation of the interventions, the CCC and Control sets remained stable and low, while Lilly's performance on the mIR set demonstrated an increasing trend. Upon the intervention + generalization condition, Lilly's flashcard performance on the mIR set remained stable and higher (range of 2 to 5 with average of 3.7 problems correct per session) than the CCC and Control sets. Lilly's performance on the Control condition was consistently low (0 or 1 problem correct for all but one session). Her performance on the CCC condition was stable and higher than the Control condition but lower than the mIR condition, with 6 consecutive intervention sessions at 2 problems correct. At the 4-week maintenance, Lilly correctly answered 5 problems in the mIR condition and 0

problems in the CCC and Control Conditions. Visual analysis across all phases of the study indicates that the mIR intervention resulted in greater improvements in flashcard performance and maintenance over time than the CCC or Control conditions.

Digits Correct Per Minute

On the daily assessment sheets, visual analysis of baseline data for DCM indicates variability for the CCC and Control conditions and stability for the mIR condition. Upon implementation of the intervention, Lilly's data indicate lower levels of performance and a decreasing trend for all three sets of problems with the sixth and final session of the intervention session at 0, 1, and 0 DCM in the mIR, CCC, and Control condition, respectively (see next paragraph for explanation). Upon the implementation of the intervention + generalization phase, Lilly's performance on the Control set remained low, while her performance on the CCC and mIR sets demonstrated an immediate increase in level. The mIR set continued on an increasing trend through the intervention + generalization phase, ranging from 3 to 8 DCM. At the 4-week maintenance, Lilly's performance on the mIR set remained higher (7 DCM) than her performance on the CCC (2 DCM) and Control (0 DCM) sets. Visual analysis across phases indicates greater performance on DCM in the mIR condition than the CCC or Control conditions, with those gains maintained at the 4-week follow-up.

Observations during the assessments indicated that during the baseline phase, Lilly quickly wrote an answer for each problem without reading the problem (i.e., guessing), which resulted in a high number of problems completed. Using the metric of DCM, Lilly received credit for many digits (e.g., 1+1 = 12 includes one correct digit) that resulted from her quick response to the problems, explaining the high rates of DCM during baseline. During the intervention, Lilly began to read each problem aloud before answering. This often took more than 5 seconds per problem, which decreased problems attempted per minute. Because of this

highly variable baseline data, the ability to draw conclusions from DCM data is limited. To correct for this guessing, percentage of problems correct per minute was calculated.

Percentage of Problems Correct

During the baseline phase, Lilly's accuracy performance on all sets was consistently low (ranging from 0% to 21% problems correct). In the intervention phase, her performance across the intervention sets remained low, with only 2 of 6 CCC assessments higher than the baseline phase and only 1 of 6 mIR sessions higher than the baseline phase. The Control condition remained below 20% during the intervention phase. In the intervention + generalization phase, Lilly's performance on the mIR and CCC conditions immediately rose and remained higher (with only one overlapping data point) than the Control condition. Performance on the mIR set of problems during the intervention + generalization phase ranged from 23% to 50% accuracy, and CCC performance ranged from 15% to 36% accuracy. At the 4-week follow-up, Lilly correctly answered 42% of problems attempted for the mIR set, 20% for the CCC set, and 0% for the Control set.

Discussion

The purpose of the current study was to compare the effectiveness of two interventions intended to increase math fact accuracy and fluency in a student with an intellectual disability. Results indicated that mIR was effective in increasing math fact accuracy and fluency as measured by flashcard automaticity, and this result was maintained at the 4-week follow-up session; however, Lilly did not improve on the paper/pencil CBMs in the mIR condition until the intervention was supplemented with an additional intervention targeting generalization. Results also showed that CCC was not an effective intervention for Lilly.

There are many possible explanations for the differential effectiveness of the mIR and CCC conditions, as there were many differences in the components that comprise the two interventions.
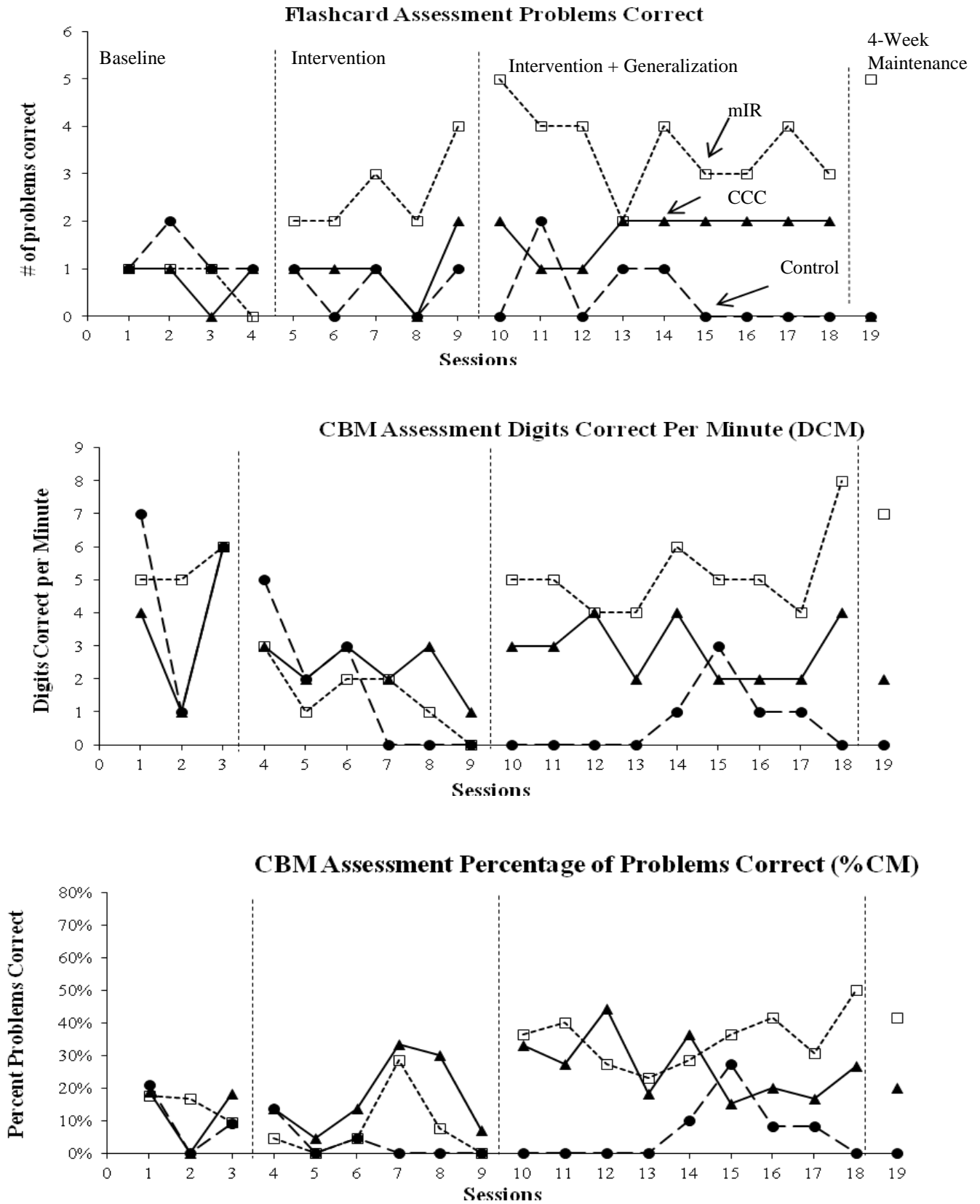
Figure 1. Number and percentage of problem/digits correct across baseline and intervention phases

One possible explanation is the difference in opportunities to respond (OTR) to target problems between the two conditions. During the initial planning of this study, the experimenters believed the CCC intervention would result in greater OTRs to target problems than the mIR intervention due to the time spent on known problems in the mIR condition; however, Lilly's slow pace on writing tasks resulted in a much greater rate of OTR in the mIR condition (which was answered verbally) than CCC (which was answered with written responses). Examination of OTR data from three random intervention sessions shows that Lilly had more than twice as many opportunities to respond to problems during the mIR condition than during the CCC condition (e.g., 53 v. 21, 68 v. 26, 73 v. 25). This is consistent with previous research indicating that interventions with more opportunities to respond result in greater learning (Cates et al., 2003). One alternative to traditional CCC is a verbal or sub-vocal CCC (Skinner, Bamberg, Smith, & Powell, 1993; Skinner, Ford, & Yunker, 1991), which may have been more effective given Lilly's slow writing skills. Future research should study the comparative effects of IR and CCC when the form of the response is the same (i.e., both responses are made verbally). This finding also has important implications for practitioners, as altering response topographies in order to provide greater OTRs may be a very effective method for improving intervention efficiency with little change to resources.

A third possible cause for the differences in effectiveness between the two interventions is that something from the incremental rehearsal procedure, outside of simply high OTRs, caused increased learning. For example, incremental rehearsal includes (1) adding target problems one at a time (i.e., incrementally), (2) increased space between newly learned items by interspersing known problems (i.e., increased time between recalls), and (3) high rates of *correct* responses due to addition of already known items, which may reduce students' frustration. It is not clear, however, whether systematically folding in new unknown problems was any more effective than a traditional drill with flashcards with 100% unknown problems. In fact, previous research has shown practice with greater ratios of unknown to known problems is more efficient than practice with lower ratios of unknown to known (Cates et al., 2003; Joseph & Nist, 2006; Roberts & Shapiro, 1996). Future researchers should examine the components of incremental rehearsal by systematically isolating parts of the intervention to find out which may be necessary and which are not. Knowing which components lead to increased learning rates may be particularly important as learning to carry out the incremental rehearsal procedure took far more time and practice from the examiners than a traditional flashcard drill, and this would make it difficult to use in certain educational situation (e.g., as a peer-mediated intervention, using parent volunteers or untrained school personnel).

One unexpected finding from this study was Lilly's difficulty generalizing from one type of intervention to a different type of assessment. On the mIR problem set, she did not generalize her performance with the flashcards to the assessment sheets until she was given extra practice focusing on generalization. On the CCC problem set, Lilly's low rates of acquisition and fluency preclude an analysis of generalization, as she would first need to acquire the skill before measuring generalization. These results support the findings of Duhon et al. (2012), who found that students' learning of math facts in one topography (computerized) did not generalize to improvements in another topography (paper/pencil assessments). The present study, coupled with Duhon et al., supports the need for researchers to measure multiple methods for demonstrating a skill and consider interventions that improve multiple methods. In addition, generalization needs to be explicitly targeted (Stokes & Baer, 1977). True proficiency in a skill requires not just skill acquisition and fluency, but utilizing the learned response with other learned responses to solve problems (Haring & Eaton, 1978). Lilly's difficulty with generalization may have been caused by difficulties with generalization across responses (writing entire problem for CCC, writing only answer for CBMs, and oral responding for flashcards) or difficulties with generalization across different stimuli (flashcards, CCC sheets, CBMs). Because generalization can take multiple forms (Kazdin, 2001), researchers targeting academic skills should consider specific types of generalization (across settings, responses, stimuli,

time, etc). In addition, this has important implications for practitioners in the schools, where the focus often remains on acquisition and fluency of a skill when it may be necessary to include additional activities that reduce OTRs but may lead to increased generalization.

Limitations

One limitation to this study relates to external validity. This was a single-case design study that took place in a university clinic, which does differ from schools. In addition, due to the method of participant recruitment, it is not possible to know how many students were recommended for this study. It is conceivable that parents who brought their children to the clinic differ in some way from the normal population. Researchers could replicate this study in schools and extend to different groups of students and different types of target skills.

A second limitation is the variability in baseline data for digits correct per minute (DCM), which limited interpretation of DCM findings. Fortunately, this study included two additional dependent variables that allowed for greater interpretation. DCM may not be an adequate dependent variable for students with very low accuracy and high rates of guessing. In addition, the participant's DCM initially decreased after beginning the intervention, when she began to read each problem rather than simply writing random numbers. Future researchers could investigate the relationship between accuracy and fluency as students move from the acquisition to the fluency-building stages of the instructional hierarchy (Haring & Eaton, 1978), particularly as it relates to how to best measure progress at each stage.

A third limitation of the study is the relatively slow growth in Lilly's math skills. Even after the intervention, her DCM remained in the frustrational range (<10DCM) for her grade (Shapiro, 2011). However, it should be considered that Lilly had finished 2nd grade with little if any math fact recall. In the course of approximately 140 minutes of intervention over 5 weeks, her accuracy on problems in the modified incremental rehearsal condition more than doubled, with improvements also noted in flashcard automaticity. Thus, growth

should be considered in the context of the student's initial skills and history of learning rate.

A fourth limitation is related to the item set sizes used for the study. Sets of 11 items were constructed and were to be targeted using flow lists. Because of Lilly's relatively slow learning rate, she did not practice all problems in the set over the course of the treatment. Having unpracticed problems on her assessment sheets may have underestimated her accuracy and fluency on targeted problems. More research is needed to understand the ideal set size, which may depend on target skill, age, and student ability.

As a final limitation, this study does not allow for an understanding of the particular aspects of the modified incremental rehearsal that led to greater improvements than the Cover, Copy, Compare intervention. As mentioned previously, future researchers should systematically analyze the components within incremental rehearsal to determine if it has added benefit beyond traditional drill and practice.

Summary

This study demonstrated the comparative effectiveness of a modified incremental rehearsal flashcard intervention as a method of improving a student's math fact accuracy and automaticity. This intervention led to greater learning rates than a CCC or control condition, though results were not generalized to different stimuli (paper/pencil) and response types (written) until additional generalization practice was included. This study has practical implications for educators in that educators must choose the most efficient intervention for a given skill while also considering generalizability to different stimuli, skills, and response topography. In addition, this study may serve as a reminder that interventions (such as Cover, Copy, Compare) that have been shown to be effective (Joseph et al., 2012) must also be continually validated with different populations and skills. Finally, both interventions should be closely examined for the effective components so that when interventions need to be adapted, the most necessary components can be retained.

References

Albers, A. E., & Greer, R. D. (1991). Is the three-term contingency trial a predicator of effective instruction. *Journal of Behavioral Education, 1*, 337-354.

Axtell, P. K., McCallum, R. S., Bell, S.M., & Poncy, B. (2009). Developing math automaticity using a classwide fluency building procedure for middle school students: A preliminary study. *Psychology in the Schools, 46*, 526-538. doi: 10.1002/pits.20395

Browder, D. M., Spooner, F., Ahlgrim-Delzell, L., Harris, A. A., & Wakeman, S. (2008). A meta-analysis on teaching mathematics to students with significant cognitive disabilities. *Exceptional Children*, *74*, 407-432.

Burns, M. K. (2005). Using incremental rehearsal to increase fluency of single-digit multiplication facts with children identified as learning disabled in mathematics computation. *Education and Treatment of Children*, *28*, 237-249.

Burns, M. K., & Boice, C. H. (2009). Comparison of the relationship between words retained and intelligence for three instructional strategies among students with below-average IQ. *School Psychology Review*, *38*, 284-292.

Burns, M. K., Kanive, R., & DeGrande, M. (2012). Effect of a computer-delivered math fact intervention as a supplemental intervention for math in third and fourth grades. *Remedial and Special Education*, *33*(3), 184-191. doi:10.1177/0741932510381652

Burns, M. K., & Sterling-Turner, H. E. (2010). Comparison of efficiency measures for academic interventions based on acquisition and maintenance. *Psychology in the Schools, 47*, 126-134. doi: 10.1002/pits.20458

Burns, M., Zaslofsky, A., Kanive, R., & Parker, D. (2012). Meta-analysis of incremental rehearsal using phi coefficients to compare single-case and group designs. *Journal of Behavioral Education*, *21*(3), 185-202. doi: 10.1007/s10864-012-9160-2

Butler, F.M., Miller, S.P., Lee, K.-h, & Pierce, T. (2001). Teaching mathematics to students with mild-to-moderate mental retardation: A review of the literature. *Mental Retardation*, *39*, 20-31.

Cates, G. L., & Rhymer, K. N. (2003). Examining the relationship between mathematics anxiety and mathematics performance: An instructional hierarchy perspective. *Journal of Behavioral Education, 12*, 23-34.

Cates G., Skinner C. H., Watson S., Meadows T., Weaver A., & Jackson, B. (2003). Instructional effectiveness and instructional efficiency as considerations for data-based decision making: An evaluation of interspersing procedures. *School Psychology Review*, *32,* 601-616.

Codding, R. S., Archer, J., & Connell, J. (2010). A systematic replication and extension of using incremental rehearsal to improve multiplication skills: An investigation of generalization. *Journal of Behavioral Education*, *19*(1), 93-105. doi:10.1007/s10864-010-9102-9

Codding, R. S., Eckert, T. L., Fanning, E., Shiyko, M., & Solomon, E. (2007). Comparing mathematics interventions: The effects of cover-copy-compare alone and combined with performance feedback on digits correct and incorrect. *Journal of Behavioral Education*, *16*(2), 125-141. doi:10.1007/s10864-006-9006-x

Codding, R. S., Hilt-Panahon, A., Panahon, C. J., & Benson, J. L. (2009). Addressing mathematics computation problems: A review of simple and moderate intensity interventions. *Education & Treatment of Children*, *32*, 279-312.

Cumming, J., & Elkins, J. (1999). Lack of automaticity in the basic addition facts as a characteristic of arithmetic learning problems and instructional needs. *Mathematical Cognition, 5*(2), 149-180.

Duhon, G. J., House, S. H., & Stinnett, T. A. (2012). Evaluating the generalization of math fact fluency gains across paper and computer performance modalities. *Journal of School Psychology, 50,* 335-345.

Haring, N. G., & Eaton, M. D. (1978). Systematic instructional procedures: An instructional hierarchy. In N.G. Haring, T. C. Lovitt, M. D. Eaton, & C. L. Hansen (Eds.), *The fourth R: Research in the classroom* (pp. 23-40). Columbus, OH: Merrill.

Ivarie, J. J. (1986). Effects of proficiency rates on later performance of a recall and writing behavior. *Remedial and Special Education, 7,* 25-30.

Joseph, L. M., Konrad, M., Cates, G., Vajcner, T. Eveleigh, E., & Fishley, K. M. (2012). Meta-analytic review of the cover-copy-compare and variations of this self-management procedure. *Psychology in the Schools, 49,* 122-136.

Joseph, L. M., & Nist, L. M. (2006). Comparing the effects of unknown-known ratios on word reading learning versus learning rates. *Journal of Behavioral Education, 15,* 69-79.

Kazdin, A. E. (2001). *Behavior modification in applied settings* (6th ed.). Belmont, CA: Wadsworth/Thomson Learning.

Kilpatrick, J., Swafford, J., & Findell, B. (2001). *Adding it up: Helping children learn mathematics.* Washington, DC: National Academies Press.

McCallum, E., Skinner, C. H., & Hutchins, H. (2004). The taped-problems intervention: Increasing division fact fluency using a low-tech self-managed time-delay intervention. *Journal of Applied School Psychology*, *20*(2), 129-147.

McLaughlin, T. F., Reiter, S. M., Mabee, W. S., & Byram, B. J. (1991). An analysis and replication of the Add-A-Word spelling program with mildly handicapped middle school students. *Journal of Behavioral Education, 1,* 413-426.

National Center for Education Statistics (2011). *The nation's report card: Mathematics 2011*(NCES 2012–458). Retrieved from the Institute of Education Sciences, U.S. Department of Education website: http://nces.ed.gov/nationsreportcard/pdf/main2011/201245 8.pdf

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. Retrieved from the U.S. Department of

Education website: http://www2.ed.gov/about/bdscomm/list/mathpanel/report/final-report.pdf.

Poncy, B. C., Skinner, C. H., & Jaspers, K. E. (2007). Evaluating and comparing interventions designed to enhance math fact accuracy and fluency: Cover, copy, and compare versus taped problems. *Journal of Behavioral Education*, *16*(1), 27-37. doi:10.1007/s10864-006-9025-7

Poncy, B. C., Skinner, C. H., & McCallum, E. (2012). A comparison of class-wide taped problems and cover, copy, and compare for enhancing mathematics fluency. *Psychology in the Schools, 49*, 744-755. doi: 10.1002/pits.21631

Poncy, B. C., Skinner, C. H., & O'Mara, T. (2006). Detect, practice, and repair: The effects of a classwide intervention on elementary students' math-fact fluency. *Journal of Evidence-Based Practices for Schools*, *7*(1), 47-68.

Roberts, M. L., & Shapiro, E. S. (1996). Effects of instructional ratios on students' reading performance in a regular education classroom. *Journal of School Psychology, 34,* 73-91.

Rhymer, K. N., Dittmer, K. I., Skinner, C. H., & Jackson, B. (2000). Effectiveness of a multi-component treatment for improving mathematics fluency. *School Psychology Quarterly*, *15,* 40–51.

Shapiro, E. S. (2011). *Academic skills problems: Direct assessment and intervention* (4th ed.). New York: The Guilford Press.

Sindelar, P., Rosenberg, M., & Wilson, R. (1985). An adapted alternating treatments design for instructional research. *Education & Treatment of Children*, *8*, 67-76.

Skinner, C. H. (2008). Theoretical and applied implications of precisely measuring learning rates. *School Psychology Review, 37,* 309-315.

Skinner, C. H., Bamberg, H. W., Smith, E. S., & Powell, S. S. (1993). Cognitive cover, copy, and compare subvocal responding to increase rates of accurate division responding. *Remedial and Special Education*, *14*(1), 49-56.

Skinner, C. H., Belfiore, P. J., Mace, H. W., Williams-Wilson, S., & Johns, G. A. (1997). Altering response topography to increase response efficiency and learning rates. *School Psychology Quarterly*, *12*, 54-64. doi:10.1037/h0088947

Skinner, C. H., Belfiore, P. J., & Watson, T. S. (2002). Assessing the relative effects of interventions in students with mild disabilities: Assessing instructional time. *Journal of Psychoeducational Assessment, 20*, 346-357. doi: 10.1177/073428290202000403

Skinner, C. H., Ford, J.M., & Yunker, B.D. (1991). A comparison of instructional response requirements on the multiplication performance of behaviorally disordered students. *Behavioral Disorders, 17*, 56-65.

Skinner, C. H., McLaughlin, T. F., & Logan, P. (1997). Cover, copy, and compare: A self-managed academic intervention effective across skills, students, and settings. *Journal of Behavioral Education*, *7,* 295-306.

Skinner, C. H., & Shapiro, E. S. (1989). A comparison of taped-words and drill intervention on reading fluency in adolescents with behavior disorders. *Education and Treatment of Children, 12,* 123-133.

Skinner, C. H., Turco, T. L., Beatty, K. L., & Rasavage, C. (1989). Cover, copy, and compare: A method for increasing multiplication performance. *School Psychology Review*, *18*, 412-420.

Stading, M., Williams, R. L., & McLaughlin, T. F. (1996). Effects of a copy, cover, and compare procedure on multiplication facts mastery with a third grade girl with learning disabilities in a home setting. *Education & Treatment of Children*, *19*, 425-434.

Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis*, *10*, 349-367.

Tucker, J. A. (1988). *Basic flashcard technique when vocabulary is the goal.* Unpublished materials.

Book Review

Book Review:
The Work-Smart Academic Planner:
Write it Down, Get it Done

Daniel F. McCleary
Stephen F. Austin State University

Dawson and Guare recently published a spiral-bound academic planner for individuals with executive skills delays entitled *The Work-Smart Academic Planner: Write It Down, Get It Done* (2015). This planner is designed for students in grades 6-12 and aligns with one of their other publications, *Coaching Students with Executive Skills Deficits* (2012). A user's guide is available at www.guilford.com/work-smart-guide. The guide provides additional tips on how to use the academic planner, including a section on troubleshooting various problems that may arise. Ideally, a student will use the planner with the assistance of an academic coach, as described in the aforementioned book. In fact, it is this author's opinion that an individual with executive skills deficits is likely to experience heightened stress levels from flipping through the planner without the guidance and support of an academic coach or knowledgeable parent. The purpose of the planner is to aid students in developing executive skills by tracking daily homework assignments and due dates, setting goals, recording accomplishments, and becoming more aware of one's own executive skills strengths and delays.

This planner is built on research-based concepts and the experienced practice of the authors at the Center for Learning and Attention Disorders at Seacoast Mental Health Center. Special educators and school psychologists may consider using the materials to help students with organization, planning, time management, and study skills. Specifically, in regard to the National Association of School Psychologists' Practice Model (NASP, 2010), school psychologists may use this as a tool to advocate for involvement in either consultation or direct service relating to the development of academic skills (Domain 3) or by providing workshop/inservice trainings on how to effectively use the planner to parents/teachers (Domain 7).

Content and Structure

The planner is separated into four different sections. The first section contains an Executive Skills Questionnaire, which identifies one's executive skills strengths and delays; Executive Skills Problem Checklist, which helps identify the student's top three executive skills impediments and how their executive skills strengths may enable them to mitigate the impediments; and an Executive Skills Tip Sheet, which provides tips on how to provide one's self with positive self-talk for each of the 11 identified executive skills areas (response inhibition, emotional control, task initiation, organization, flexibility, goal-directed persistence, working memory, sustained attention, planning/prioritization, time management, and metacognition).

Section two focuses on both long-term and short-term goal setting. A Task Completion Checklist allows the individual to keep track of personal goals, such as plans after high school, academic goals for the year and marking periods, and goals for individual class grades. A Goal Setting Worksheet provides graphic organizers and prompts for setting each of the goals contained in

the Task Completion Checklist. In essence, the Goal Setting Worksheet breaks the Task Completion Checklist into a guided task-analysis exercise.

The third section provides monthly and daily plans. In all, 48 weekly and 12 monthly planners are provided. The monthly planner aids students in recording long-term assignments and goals; whereas, the daily planner is for tracking short-term assignments and goals. It is recommended that the monthly planner is checked daily to ensure that important due dates are not forgotten. At the top of each monthly planner are sections to record one's executive skills goal and academic goal for the month. The daily planner includes a place to record one's long-term goal at the top of the page. On the daily planner, each day includes places to record urgent information, subject, due date, how long the task is expected to take, anticipated start time on the task, materials that are needed to complete the task, and a place for contact information.

The final section emphasizes various strategies one can use to increase one's chance of success. For example, strategies for studying for tests, managing distractions, managing time, and increasing reading comprehension are provided along with a four day study plan and graphic organizer, a five-paragraph essay template, and a long-term-project planning form.

Critique

The planner provides good task analysis, tips, and strategies to both short- and long-term goals, which separates it from any typical planner. Regardless, it is strongly recommended that the planner be used in conjunction with an academic coach, especially if the student experiences significant executive skills delays. Despite the positives, there are some drawbacks. The cover and size may be embarrassing to some adolescents as it is visually different from what their peers are likely to be using. Also, the student must write the numeral dates into the planner. While this is likely to be a tedious task, emphasizing again the importance of an academic coach, it also provides flexibility in that the planner is not bound to one specific academic year.

In sum, this academic planner could improve the lives of individuals struggling with executive skills deficits. It is most likely to be effective when paired with an academic coach. As an alternative to individual academic coaches, a school may opt to purchase the planners for students with special needs and provide regular instruction and assistance with the planner as part of their daily activities under the direction of a special education teacher or school psychologist.

References

Dawson, P., & Guare, R. (2012). *Coaching students with executive skills deficits*. New York: Guilford.

Dawson, P., & Guare, R. (2015). *The work-smart academic planner: Write it down, get it done*. New York: Guilford.

National Association of School Psychologists. (2010). *Model for comprehensive and integrated school psychological services*. Retrieved from http://www.nasponline.org/standards/practice-model/10_domains_of_practice.pdf