Please use the following citation when referencing this work:

La Spata, M. G., Carter, C. W., Johnson, W. L., & McGill, R. J. (2015). Evaluating video self-modeling treatment outcomes: Differentiating between statistically and clinically significant change. *Contemporary School Psychology, 20,* 170-182. doi: 10.1007/s40688-015-0072-8

The final publication is available at Springer via http://dx.doi.org/10.1007/s40688-015-0072-8.

# Evaluating Video Self-Modeling Treatment Outcomes: Differentiating between Statistically and Clinically Significant Change

Michelle G. La Spata

Christopher W. Carter

Wendi L. Johnson

Ryan J. McGill

Texas Woman's University

Author Note

Correspondence concerning this article should be addressed to Michelle G. La Spata, Department of Psychology and Philosophy, Texas Woman's University, P. O. Box 425470, Denton, TX 76204. E-Mail: mlaspata@twu.edu

**Abstract**

The present study examined the utility of video self-modeling (VSM) for reducing externalizing behaviors (e.g., aggression, following directions, hyperactivity, and impulsivity) observed within the classroom environment. After identification of relevant target behaviors, VSM interventions were developed for first and second grade students ($N = 4$), ages 7-8, from a suburban school district. All participants were eligible for special education and related services, with eligibility classifications that ranged from Other Health Impairment (OHI) to Traumatic Brain Injury (TBI). VSM treatment outcomes were evaluated using multiple change indices (e.g., Reliable Change Index, Means Difference Effect Size) to examine the clinical and statistical significance of differences observed between pre and post intervention administrations of parent and teacher forms of the Behavior Assessment System for Children-Second Edition (BASC-2) rating scale. Although outcomes for individual participants were relatively consistent across the multiple change indices, differential effects in reductions in reports of externalizing behaviors across teacher and parent rating forms were observed. Implications, limitations, and future directions of VSM research are also discussed.

*Key words*: video self-modeling, externalizing behaviors, RCI, treatment outcomes

**Evaluating Video Self-Modeling Treatment Outcomes: Differentiating between Statistically and Clinically Significant Change**

Videotaped self-modeling in which individuals are shown examples of themselves performing a target behavior accurately are an intervention that has been shown to be an effective treatment for a variety of behaviors across ages and settings. Within the technical literature, positive results have been demonstrated for treating or teaching behaviors ranging from math skills (Schunk & Hansen, 1989) to depression (Kahn, Kehle, Jenson, & Clark, 1990). Despite its versatile clinical utility, video self-modeling is most often utilized to promote adaptive social skills and related functioning with children and adolescents within the school environment (e.g., Bilias-Lolis, Chafouleas, Kehle, & Bray, 2012; Buggey & Ogle, 2012; Burton, Anderson, Prater, & Dyches, 2013; Gelbar, Anderson, McCarthy, & Buggey, 2012; Kehle, Madaus, Baratta, & Bray, 1998; Ohtake, Takeuchi, & Watanabe, 2014; Possell, Kehle, McLoughlin, & Bray, 1999).

Video self-modeling (VSM) is rooted in social learning theory (e.g., Bandura, 1986), which suggests that some behaviors are learned and reinforced vicariously through models or cues in the social environment. While adults typically serve as models for younger children, peers begin to influence behavioral performance as children progress through natural developmental stages (Dishion, Piehler, & Myers, 2008; Dodge & Godwin, 2013). Modeling is diverse and generally refers to whenever an adult or peer demonstrates behaviors that can serve as an example of how to perform correctly within the social environment (Buggey & Ogle, 2012). Bandura (1986) suggests that model-viewer congruence can influence the degree to which the viewer will ultimately attend to the model, accordingly, VSM emphasizes using the individual as their own model (via video) to strengthen learning and generalization of more adaptive behaviors.

VSM involves the student watching and learning from his or her own positive behavior through a video medium (Rudy et al., 2014). Positive behaviors or target skills are elicited from the students through scaffolding supports such as direct instruction, peer or adult modeling, or role-playing, and then captured on video. After the video is edited to depict only the adaptive behavior with the prompts/supports edited out, the student watches their own individualized movie (Collier-Meek et al., 2011). In spite of past research documenting its effectiveness (e.g., Bellini & Akullian, 2007; Hithcock, Dowrick, & Prater, 2003; Mason et al., 2012), VSM has not been widely implemented within the educational environment by school-based practitioners (Bray & Kehle, 2012). Barriers to implementation may include negative perceptions regarding the logistics behind video self-modeling, including the extent of technological knowledge and time required to plan, record, edit, and show the videos (Wilson, 2013). Additionally, the materials required may not be readily available in some schools (Buggey, 2007), although many of these perceived barriers have been alleviated recently with increased availability of digital video technology on phones, tablets, and laptops (Collier-Meek et al., 2011).

**Clinical Applications of VSM**

As a result of the proliferation of video modeling technology research on the potential clinical applications of VSM has intensified over the course of the last decade. The vast majority of this research has examined the use of VSM for increasing the fluency of targeted social skills in students diagnosed with autism spectrum disorder (e.g., Bellini & Akullian, 2007; Gelbar et al., 2012). Buggey and Ogle (2012) reported fourteen self-modeling studies that were conducted using students with autism spectrum disorder, including a meta-analysis of 23 single-subject designs conducted by Bellini and Akullian (2007). Results of these analyses suggest that VSM was an effective intervention when used to promote adaptive social behaviors across a variety of

settings. Additional research on VSM has addressed externalizing behaviors which included behaviors such as fighting, inappropriate social interactions with peers and adults, touching, vocalizing, aggression, and making inappropriate noises (Madaus & Ruberto, 2012). As an example, Buggey (2005) found that VSM was effective in reducing school-based tantrums in two male participants diagnosed with autism spectrum disorder. The researchers identified the triggers for the tantrums using a functional behavioral assessment, and those triggers were then targeted in the scenes in the video (Gelbar et al., 2012). Similarly, Bilias-Lolis et al. (2012) found that video self-modeling was effective in reducing disruptive behaviors in three adolescent students with intellectual disabilities.

More inconsistent results have been found in examinations of the efficacy of VSM on reducing inappropriate or disruptive behaviors among students with emotional and behavioral disorders. Whereas some researchers (e.g., Davis, 1979; Kehle, Clark, Jenson, & Wampold, 1986; McCurdy & Shaprio, 1988) have found VSM to be effective in reducing inappropriate behaviors across sample sizes ranging from one to four participants, other studies have found VSM to be less effective. For example, Shear and Shapiro (1993) implemented a VSM intervention with six students, ages 7 to 12 years old, who attended a school specifically designed to address needs related to emotional disturbance. The student's baseline levels of disruptive behavior were measured via observing and videotaping them while in the classroom. After the initial footage was obtained, the tapes were edited to remove all instances of disruptive behavior so that the remaining footage consisted of the students engaging in appropriate behavior(s), with the final tapes lasting approximately 15 minutes apiece. Additionally, half of the students also participated in a condition where they were trained to self-record their appropriate and inappropriate behaviors. It was found that the students did not demonstrate

significant improvements in their behavior, regardless of whether they self-recorded or not. It was suggested that video length may have moderated the effectiveness of the intervention protocol.

Clark et al. (1993) also found no significant improvements in behavior using VSM with a sample of male preschool students ($N = 6$) diagnosed with oppositional defiant disorder. Each student participated in three conditions which included: self-modeling, peer modeling, and a control condition. Although an attempt was made to utilize shorter intervention tapes, the length remained approximately 4-5 minutes in duration. Results were attributed to the participant's ages and level of cognitive development. Specifically, it was noted that although preschoolers could observe and imitate the behaviors of others, the cognitive processes (e.g, attentional control) required to concentrate during modeling procedures of that length are not fully developed at that point of time in the age span.

Possell et al. (1999) also found differential outcomes in reducing inappropriate behaviors among students with emotional disturbance. Participants consisted of four male students (ages 5-8 years old), receiving instruction in general education and self-contained classrooms. In contrast with aforementioned VSM research, these students were actively involved in selecting behaviors to be edited within their videos. Whereas results indicated positive changes in behavior among two of the older students in the sample, the intervention was not as effective with the younger participants. Again, it was suggested that developmental variables (e.g., cognitive development, self-appraisal, self-regulation skills, and ability to use predictive forethought) may have influenced the efficacy of that particular application of VSM intervention.

A review of previous studies highlighting ineffective applications of VSM for reducing disruptive behaviors in students with behavioral and emotional disorders reveals several salient

findings. First, the vast majority of participants did not appear to be actively involved in the video making process. In other words, the interventionists recorded all of their behaviors in the classroom setting and then edited the tapes to retain only the desirable behaviors. Second, interventionists appeared to focus on the appropriate behaviors exhibited by the individual students, as there was no mention of taping any peer interactions or other group situations (e.g., Hitchcock, Dowrick, & Prater, 2003). Third, interventionists did not appear to elicit information related to potential target behaviors through validated procedures such parent and teacher interviews. Finally, most of the aforementioned studies relied exclusively on single-case design methodologies to evaluate the efficacy of VSM interventions. As a result, salient examinations of VSM outcomes using more robust methods for evaluating treatment efficacy (e.g., Lambert, Hansen, & Bauer, 2007) are currently lacking within the technical and empirical literature.

**Purpose**

The purpose of the current investigation is to examine the potential impact of involving students in the VSM intervention process for reducing externalizing behavior problems. According to Buggey (2007), children who exhibit behavior difficulties can perform correct responses to social situations during role-playing activities. Furthermore, it is suggested (e.g., Buggey, 2007) that they may be more enthusiastic when they are actively involved in the role-playing activity and are able to participate in the videotaping process, potentially leading to more positive intervention outcomes. Additionally, the current study utilized scripts, included in role-playing situations with peers participating in the VSM intervention, that take into account potential functions of the identified problem behaviors. Thus, in contrast to previous research, the participants in the current study were actively involved in the video-making process.

Additionally, more research is needed in order to determine if positive intervention outcomes can be obtained using shorter intervention tapes (e.g., less than 5 minutes). Accordingly, the current study utilized videos of approximately 2 minutes with an early elementary age sample to examine whether shorter videos could be effective with this age group in reducing reports of aggressive and other related externalizing behaviors. In sum, our intention is to examine combinations of variables that potentially contribute to the efficacy of VSM with early elementary-aged children. It is believed that the current investigation will contribute to the expanding school-based intervention literature generally and will extend our understanding of potential moderating variables that influence the effectiveness of VSM interventions in clinical settings.

## Method

### Participants

The demographic characteristics of study participants are provided in Table 1. Four elementary school students from a suburban school district were recruited for the study. These students were 7-8 years old and consisted of three male students and one female student. Qualifying criteria for children to participate in the study were that they be eligible for special education and related services and had known difficulties with externalizing behaviors (e.g., aggression, conduct problems, and/or hyperactivity), observed within the educational setting. Accordingly, all research participants received special education and related services for more than half of the school day in a special education classroom designed to meet emotional and behavioral needs. Parent consent and child assent were obtained for all participants. Additionally, all procedures described below were sanctioned and approved by the Institutional Review Board (IRB) at Texas Woman's University to ensure the welfare of participants.

**Assessment Procedures**

Pre-intervention assessment, consisting of parent and teacher interviews, and student observations, was conducted for each participant to facilitate the development of an individualized VSM intervention designed to address identified target behaviors.

Participant 1 (John) is an eight-year-old, Caucasian second grade male who was eligible for special education services under the labels of emotional disturbance (ED) and other health impairment (OHI) due to his medical diagnosis of attention deficit hyperactivity disorder (ADHD). The parent and teacher both noted that John exhibited aggression due to impatience with peers and transitioning from preferred activities. The physical aggression that he exhibited consisted of punching, kicking, shoving, or biting others. Therefore, John's intervention was developed for the purpose of reducing aggressive behaviors.

Participant 2 (Steven) is an eight-year-old, Caucasian second grade male who was eligible for special education services under the criterion of ED and speech impairment (SI) due to articulation. Based on parent and teacher interviews, Steven appeared to exhibit physically aggressive behaviors towards staff and peers when others did not comply with his directives, or when others received more attention than he was from staff members during classroom observations. These physically aggressive behaviors manifested as pushing, shoving, kicking, or destroying the property of others. Accordingly, Steven's VSM intervention was developed in order to also reduce aggressive behaviors.

Participant 3 (Jose) is a seven-year-old, Hispanic first grade male who was eligible for special education services for OHI due to ADHD. Both the parent and teacher reported that their primary concerns were that Jose did not stay on task when he was doing his work. In addition, his teacher expressed concerns with his voice volume, as he tended to yell when he spoke. Jose

exhibited hyperactivity and impulsivity and was easily distracted by peers during classroom observations. Although he usually did not manifest aggressive behavior spontaneously, he did behave more aggressively in response to the behavior of his peers. The intervention target for Jose was to increase on task behavior as well as reduce hyperactivity and impulsivity.

Participant 4 (Kathy) is a seven-year-old, Caucasian first grade female who was eligible for special education services under the categories of OHI due to traumatic brain injury, and SI due to articulation and language difficulties. Based on parent and teacher interviews, Kathy did not present with physically aggressive behaviors, however she would often refuse to follow directions and expressed frustration when the teacher provided feedback regarding inappropriate behavior during observations. Kathy's VSM intervention was developed to increase the number of times she followed directions when requested without engaging in relational aggression.

**Outcome Measure**

Externalizing behaviors were assessed using parent and teacher forms of the Behavior Assessment System for Children-Second Edition (BASC-2; Reynold & Kamphaus, 2004a), which provided both the measure of pre-post behavioral change as well as the criterion measure for classification of participants into outcome groups (e.g., Jacobson & Truax, 1991). The BASC-2 is a standardized rating scale that is frequently utilized within school-based settings to measure child and adolescent behavior. The norming samples for the parent and teacher forms ranged from 4,600 to 4,800 participants and is stratified according to region, community type, sex, and race, and is nationally representative based upon 2001 U.S. census estimates. Extensive normative and psychometric data can be found in the BASC-2 Technical Manuals (Reynolds & Kamphaus, 2004b). Average internal consistency estimates for the included scores in this study ranged from .92 to .97 across standardization and clinical samples whereas test-retest coefficients

ranged from .90 to .91. Validity evidence was provided in several forms in the Technical Manual and independent reviews are available (Stein & Watson, 2007).

In this study, BASC-2 rating forms were administered to parents (PRS) and teachers (TRS) of study participants prior to the beginning of VSM treatment to establish baseline levels of behavioral performance and post-intervention later to examine the effectiveness of intervention outcomes. BASC-2 child age parent and teacher forms (ages 6-11) contain 139 and 160 questions respectively. Each item on the BASC-2 is rated using a four-point Likert scale (i.e., *Never*, *Sometimes*, *Often*, *Almost Always*) that asks raters to evaluate the degree to which they have observed reference behaviors within their respective environments. These ratings are linearly combined to form a multitude of internalizing, externalizing, and adaptive behavior scale and composite indices expressed as T-scores ($M = 50$, $SD = 10$). Interpretative guidelines provided within the Technical Manual (Reynolds & Kamphaus, 2004b) are as follows: T-scores ranging from 40 to 59 are within the average range; T-scores ranging from 60 to 69 are within the "at-risk" range, indicating a potential problem that needs monitoring, and T-scores above 70 are within the "clinically significant" range, indicating a serious concern that warrants immediate attention. The focus of the current investigation was on the significance of change observed across the BASC-2 Externalizing Problems composite score. The Externalizing Problems composite is a linear aggregate of several subscales associated with aforementioned participant target behaviors (e.g., Hyperactivity, Aggression, and Conduct Problems).

**Intervention and Data Collection Procedures**

Parents, teachers, and students were interviewed separately using protocols specifically designed for the current investigation. Interview protocols were designed to elicit relevant information regarding potential target behaviors as well as variables impacting behavioral

performance across settings. Additionally, two classroom observations were conducted for each participant to validate information obtained in interviews. All interviews, observations, and rating forms were administered and completed by two student researchers with advanced training in the application of VSM methodologies in school-based settings under the supervision of the lead author.

As noted previously, the investigators established an individualized behavioral target that served as the focus on for the VSM intervention based on these data obtained from the pre-intervention assessments. VSM interventions were delivered to each student over six concurrent sessions. The first session consisted of rapport-building activities in order to allow the students a chance to become familiar with the investigators. This was done typically in the form of playing a board game or doing a similar activity with each child. The second session was a mock run-through of the VSM scenarios, which involved the child completing a skit or role-playing a social situation that followed their target behavior while a camera, which was not turned on, was facing the child to have them become familiar with the camera's presence. The third and fourth sessions consisted of filming the VSM stories which took place in the school setting. Afterwards, the video was then edited to demonstrate mastery of the target behavior. During the fifth session, the investigators introduced the individualized movie to each student and walked through the meaning of the video. The student then watched the video for 10 consecutive school days. Treatment integrity (e.g., watching the intervention videos) was documented by data collection completed by each participant's classroom teacher. 100% fidelity was observed across all participants. At the conclusion of the intervention period, researchers again administered rating scales to parents and teachers in order to provide a standardized assessment of client change.

**Data Analysis**

Multiple methods, such as the Reliable Change Index, have been proposed within the technical literature over the last two decades for evaluating psychosocial change (see Lambert, Hansen, & Bauer, 2007). Despite the increase in the number of indicators now available for practitioners to utilize when appraising these data, questions have been raised about the consistency of information provided by multiple outcome indicators. In a recent comparison of multiple statistical techniques for evaluating change Bauer, Lambert, Nielsen (2010), found that treatment outcomes for individual patients was moderated by the particular method utilized by researchers to evaluate pre-post change. Researchers (e.g., Busse, McGill, & Kennedy, 2014; McGill & Busse, 2014) have recently encouraged practitioners to utilize multiple outcome indicators when conducting treatment outcome evaluations to account for the potential artifacts introduced by such method variance. Accordingly, the effectiveness of VSM in reducing reports of externalizing behavior problems was estimated using several validated procedures for examining the statistical and clinical significance of change observed across the administration of pre and post intervention measures. Individual indicators are described in more detail below.

**Jacobson-Truax Method.** Jacobson and Truax (1991) initially proposed the application of a two-step procedure for assessing the clinical significance of client change which they termed the *Reliable Change Index* (RCI). Step one of the Jacobson-Truax method (JT) is to determine whether the observed change is statistically reliable, that is whether observed difference scores can be attributed to legitimate intervention effects or measurement error contained within the outcome measure using the RCI formula first articulated by Nunally and Kotsche (1983) and later expanded by Jacobson and colleagues (1984). The JT RCI is calculated by dividing the difference between pretest and posttest scores by the standard error of the differences (see Table

2.1 and 2.2). RCI values are determine to be statistically significant if they are lower than -1.96 (i.e., $p \leq .05$). The RCI is an effective statistic for use with small sample sizes due to its focus on change in the individual (Zahra & Hedge, 2010). Schmitt et al. (2013) reviewed several studies that utilized the RCI, including those that compared the RCI with several other statistical methods (e.g., Marsden et al., 2011), and found that the RCI was an appropriate method for evaluating the effects of medical, psychotherapeutic, and neuropsychological interventions.

The second step in the JT method is to determine the clinical significance of change by estimating a relevant cutoff point between dysfunctional and functional populations. According to Jacobson et al. (1984), significant change is evidenced when an individual moves from one population to the other as a result of having been provided a psychological intervention. Although several cutoffs have been proposed within the literature, a more in depth discussion of the strengths and weaknesses of each approach is beyond the scope of the current investigation. In order to provide an appropriate balance between Type I and Type II error, the current study utilized two cutoffs to evaluate the clinical significance of change as recommended by Lunnen & Ogles (1998). The first cutoff, which we refer to as the *Clinical Significance Criteria*, reflects 2 standard deviations below the established BASC-2 pretest group mean as the reference cutoff, also known within JT nomenclature as *Cutoff A*. According to Jacobson and Truax (1991), *Cutoff A* is preferred when no control group data are available as is the case in the present study. Additionally, in clinical settings *Cutoff A* has been shown to have high sensitivity, that is, it is highly unlikely that scores below this threshold represent false negatives (Bauer et al., 2004). The second cutoff, which we refer to as the *Nominal Significance Criteria*, denotes posttest scores equal to or less than 59 which corresponds to the threshold for BASC-2 scores associated with “average” or normalized functioning within the reference population. We believe that the

Nominal Criteria provides a useful contrast to the more stringent Clinical Criteria and is potentially more useful for detecting meaningful change in short-term school-based interventions such as those typically employed by school psychologists. The JT method classifies individuals as having *Recovered* (passed both criteria), *Improved* (passed RCI but not cutoff criterion), *Unchanged* (passed neither criterion), or *Deteriorated*.

**Gullikson-Lord-Novick Method.** Although the JT method for assessing clinically significant change is one of the most frequently reported by researchers, it has been challenged by several researchers (e.g., Hsu, 1989; Speer, 1992) on the grounds that the formula fails to take into account regression to the mean. To account for potential regression artifacts we calculated alternative RCI change scores using the Gullikson-Lord-Novick Method (GLN; Hsu, 1999). The GLN formula differs from the JT formula it takes into account deviation of observed scores from reference means. The GLN formula (Table 2.3) attempts to control for potential confound by including a hypothesized population mean toward which the observed scores would regress. Although more robust RCI formulas (e.g., Bryk & Raudenbach, 1992; Hageman & Arrindell, 1999) have been proposed more recently in the empirical literature, statistical assumptions for their use precluded their inclusion in the present analyses. Consistent with the interpretive guidelines for regression-based RCI indices offered by Speer (1992), a slightly more conservative critical value (i.e., $\leq$ -2.00) for detecting statistically reliable change via the GLN formula was adopted in the present study.

**Means Difference Effect Size.** To estimate the magnitude of change between pretest and posttest measures a means difference effect size (Table 2.4) was calculated using a derivation of the Hedges' g formula (Hedges & Olkin, 1985) that utilizes a pooled deviation metric (Table 2.5) as the divisor that takes into account sample size. Researchers (e.g., Durlak, 1999; Thompson,

1999) have long encouraged the reporting of effect size indicators in applied psychological research as an alternative to traditional significance testing. Effect size measures provide a useful alternative to RCI indices because they have established parameters thus allowing for interval estimation and the more precise calculation of confidence intervals (Cumming & Finch, 2001), and they permit the assessment of group level effects which are important for examining the potential effectiveness of larger scale interventions implemented by school psychologists (Volker, 2006). Guidelines for interpreting means difference effect sizes are found in Cohen (1988); they are "small," .02; "medium," .05; and "large," .08.

## Results

Descriptive statistics, including pre and post intervention BASC-2 Externalizing Problems composite T-scores and group pre and posttest means and standard deviations, are provided in Table 3. Individual outcomes measured across multiple change indices and cutoff criteria are provided in Table 4.

### BASC-2 Teacher Rating Scale

Overall, classroom teachers for the intervention participants as a whole observed improvements related to the frequency of externalizing behaviors as reflected in a decline of post intervention BASC-2 scores ($M_{TRS\ POST}$= 67.75, *SD* = 6.50) from pre intervention levels ($M_{TRS\ PRE}$= 77.25, *SD* = 8.14). Individual difference scores ranged from -12 (John) to -5 (Kathy), indicating fairly consistent declines across the group as a whole. The magnitude of change reflected at the group-level corresponded to a large effect size estimate (g = 1.12, 95% CI [-3.98, 6.23].

Using the conventional JT method, RCI values ranged from -4.90 (John) to -2.04 (Kathy). Thus, all of these JT values exceeded the threshold for statistically significant change at

the 95% confidence level (-1.96). GLN results, that took into account potential measurement artifacts associated with regression to the mean, were slightly more conservative ranging from -2.56 (John) to -0.85 (Kathy). All of the GLN values exceeded the threshold for statistically significant change at the 95% confidence level (-2.00) with the exception of Kathy (GLN = -0.85). Additionally, the distribution of TRS scores is plotted graphically in relationship to pre-established cutoff criteria (i.e., clinical versus nominal) in Figure 1. Inspection of these results reveals that none of the participant's outcome results can be interpreted as clinically significant based upon established interpretive guidelines (e.g. Jacobson & Truax, 1991). Consistent with the RCI values, MDES indicators were consistently robust ranging from 1.42 (John) to 0.59 (Kathy), with 75% of the participants demonstrating large intervention effects.

To compare differences in classification rates across the three outcome measures, Kendall's coefficient of concordance statistic (*W*) was computed. This omnibus test follows a chi-square distribution (Legendre, 2005). *W* was not significant across the three methods (*W* = .91), $\chi^2$ (3) = 8.20, *p* = .04, suggesting that overall the methods do not differ significantly from one another in classifying the rate of reliable change. Nevertheless, a synthesis of VSM outcomes reveals that while no participants recovered (e.g., exceeded both statistical and clinical thresholds for change), all participants improved with the exception of Kathy (unchanged).

**BASC-2 Parent Rating Scale**

As previously discussed, BASC-2 rating scales were also administered to the mothers of intervention participants in order to assess potential generalization effects of the individual VSM interventions. Overall, parents for the intervention participants as a whole observed improvements related to the frequency of externalizing behaviors as reflected in a decline of post intervention BASC-2 scores ($M_{PRS\ POST}$= 58.50, *SD* = 4.12) from pre intervention levels

($M_{PRS\ PRE}$= 63.275, *SD* = 9.46). Individual difference scores ranged from -22 (John) to +5 (Jose), indicating significant variation in outcomes across participants. Accordingly, the magnitude of change reflected at the group-level corresponded to a moderate effect size estimate (g = 0.63, 95% CI [-4.43, 5.68].

Using the conventional JT method, RCI values ranged from -5.88 (John) to 0.80 (Jose). Due to the increased amount of variability in the PRS ratings, only John's JT value (-5.88) exceeded the threshold for statistically significant change at the 95% confidence level (-1.96). GLN results ranged from -4.97 (John) to 0.93 (Jose). Consistent with JT results, only John's GLN value (-4.97) exceeded the threshold for statistically significant change at the 95% confidence level (-2.00). Similar to TRS outcomes, the distribution of PRS scores are plotted graphically in relationship to pre-established cutoff criteria (i.e., clinical versus nominal) in Figure 2. Inspection of these results reveals that although none of the participant's outcome results can be interpreted as clinically significant based upon established interpretive guidelines (e.g. Jacobson & Truax, 1991), Jose and John's post intervention scores fell within the average range of functioning indicating that their intervention outcomes may be interpreted as reflecting nominal significance. However, it should be noted that Jose's baseline score was also in the average range therefore the significance of his intervention results should be interpreted with caution. MDES indicators were also inconsistent ranging from 2.62 (John) to -0.34 (Kathy), with only John demonstrating large effects.

To compare differences in classification rates across the three outcome measures, Kendall's coefficient of concordance statistic (*W*) was not significant across the three methods (*W* = .91), $\chi^2$ (3) = 8.20, *p* = .04, suggesting that overall the methods do not differ significantly from one another in classifying the rate of reliable change. Nevertheless, a synthesis of VSM

outcomes reveals that John's results can be interpreted as evidence of recovery whereas the remaining participants results provide limited evidence of improvement.

## Discussion

This study examined the effects of VSM on reducing externalizing behaviors with early elementary school participants. Intervention outcomes were measured using the Externalizing Problems composite on the BASC-2 parent and teacher response forms. Consistent with best practice recommendations (e.g., Busse, McGill, & Kennedy, 2014; McGill & Busse, 2014) multiple outcome indicators were utilized to examine the statistical and clinical significance of change results.

The results of the current study indicate that our hypothesis that the combination of student involvement and shorter video length in the VSM intervention would reduce the number of aggressive and externalizing behaviors reported for early elementary-aged children with behavioral disorders was supported in part. Based on the information provided by the teacher and the parents on the BASC-2, most of the participant's demonstrated decreases in reported problem behaviors however the magnitude of these reductions consistently exceeded established thresholds (e.g., Jacobson & Truax, 1991) for clinical or statistical significance only in the school setting as reflected by the BASC-2 TRS ratings. With the exception of John, the positive TRS ratings did not generalize to the home setting as reflected in the BASC-2 PRS ratings. These inconsistent results comport with previous research examining the efficacy of VSM technologies with individuals diagnosed with emotional and behavioral disorders (e.g., Clark et al., 1993; Mason et al., 2012; Possell et al., 1999). Although the overall implication of these findings should be interpreted cautiously given the relatively small sample sizes that have been employed, it is certainly possible that younger children with behavioral and emotional disorders may have

difficulty processing information presented through visual technologies such as VSM in a way that facilitates generalization across settings. Nevertheless, the positive TRS outcomes suggest that previous non-positive VSM interventions with younger participants (e.g., Sheer & Shapiro, 1993) may have been mediated by use of longer intervention tapes (e.g., exceeding 10 minutes) as the present study utilized tapes of significantly shorter length.

As previously mentioned, while reductions in externalizing behaviors were relatively consistent across participants within the academic setting, these singificant results did not appear to generalize to the home environment for most participants. We believe there are several reasons for this finding. First, the intervention was implemented while the student's were at school and the video clips were set in the school environment. Secondly, with the exception of John, whose mother rated him within the clinically significant range on the BASC-2 PRS, teacher ratings were consistently higher on the TRS than the parent endorsements at baseline which potentially created a ceiling effect that was a measurement artifact at the onset of the intervention. Although all participants had the same teacher, they all had different parents who likely had varying levels of tolerance for externalizing behavior problems. The third reason may be that the behaviors manifested in the school environment may not be consistent with the behaviors exhibited at home. As a result, additional research is needed to examine the utility of VSM interventions as it relates to generalization across settings.

Finally, it should be noted that the relatively high concordance of results across multiple outcome indicators across all intervention participants suggests that the omnibus recommendation (e.g., McGill & Busse, 2014) for school psychologists to routinely utilize multiple outcome indicators when examining treatment effects for school-based interventions may be overstated. In the present analysis, participants maintained their same rank order across

multiple variations of the RCI (e.g., simple difference versus regression). It is for this reason that some researchers (e.g., McGlinchey, Atkins, & Jacobson, 2002) have encouraged practitioners to forego more esoteric applications of the RCI in favor of the conventional JT method. However, non-trivial differences were noted between quantitative indicators and the methods employed to appraise the clinical significance of change. Whereas moderate to large TRS intervention effects were demonstrated consistently for participants across multiple outcome indicators, the magnitude of these changes did not exceed *a priori* cutoffs (e.g., Jacobson & Truax, 1991), suggesting that the choice between utilizing a quantitative estimate of reliable change and procedures that attempt to categorize the significance of the observed change may result in different conclusions about the effectiveness of an intervention. Additional research examining the potential benefit of employing clinical significance methods for individual decision-making in school psychology is needed. Until such research is provided, we encourage practitioners to be mindful of potential threats that are introduced by transforming continuous data into nominal categories (e.g., MacCallum, Zhang, Preacher, & Rucker, 2002).

**Social Validity**

Although consistently non-significant quantitative changes were not noted at home for most participants, posttest qualitative information was gathered based on parent interview data in which parents of all four students noted an increase in prosocial behavior in their children. Therefore the parents endorsed positive changes in behavior with their children, even though this was not generally evidenced in the BASC-2 PRS outcome data. Three of the four parents also stated that their children enjoyed watching their VSM videos and looked forward to viewing them. This supports the concept that students may view VSM as an enjoyable intervention although the potential relationship between client motivation and VSM outcomes is not yet

understood. The discordance in the present study between verbal reports of perceived intervention effectiveness and the results of more systematic analyses of intervention results suggests that practitioners should use in caution when relying exclusively on social validation methods for determining the efficacy of interventions in applied settings (Foster & Mash, 1999).

**Limitations**

This study is of course not without limitations. Most notably, this study included a relatively small number of participants in the VSM protocol. However, we believe that such sample sizes more realistically reflect the day to day practice of school psychologists who routinely implement interventions within school and/or other clinical settings. In addition, the current design did not factor in potential confound effects from other interventions that took place concurrently within the classroom environment. In the school environment, it is not uncommon for multiple positive behavior interventions, such as a token economy system or differential reinforcement, to be implemented simultaneously within the classroom setting. Finally, it should be noted that the BASC-2 parent and teacher forms are an indirect measure of the externalizing behaviors; nevertheless, the utilization of a standardized measure allowed for the calculation of relevant change indices and provided a consistent protocol for measuring intervention effects.

**Conclusion**

Overall, the results of the current study support the use of VSM as a behavioral intervention for decreasing externalizing behaviors in the school environment with early elementary age students as well as with individuals with known behavioral disorders (e.g., emotional disturbance). Participant's evidenced statistically significant decreases in externalizing behaviors on the BASC-2 TRS after the intervention was implemented across the course of the

school year across multiple outcome measures. As a consequence, the use of shorter intervention tapes appeared to mediate outcomes with younger participants within the school setting. Nevertheless, these results failed to generalize to the home setting as reflected in PRS ratings and did not exceed established thresholds for clinical significance. Therefore, future studies in examining the efficacy of VSM will need to focus on identifying the variables and mechanisms responsible for positive effects with participants so that the intervention can be applied more successfully to other clinical populations and settings.

## References

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

Bauer, S., Lambert, M. J., & Nielsen, S. L. (2010). Clinical significance methods: A comparison of statistical techniques. *Journal of Personality Assessment, 82,* 60-70. Doi: 10.1207/s15327752jpa8201-11

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

Bilias-Lolis, E., Chafouleas, S. M., Kehle, T. J., & 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

Bray, M.A., & Kehle, T. J. (2012). Introduction to the special issue self-modeling: Self-modeling as a treatment for a myriad of issues. *Psychology in the Schools, 49,* 1-2. doi: 10.1002/pits.20620

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.

Buggey, T. (2005). Video self-modeling applications with children with autism spectrum disorder in a small private school. *Focus on Autism and Other Developmental Disabilities, 20,* 52-63. doi: 10.1177/10883576050200010501

Buggey, T. (2007). A picture is worth…video self-modeling applications at school and home. *Journal of Positive Behavior Interventions, 9,* 151-158. doi: 10.1177/10983007070090030301

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

Burton, C. E., Anderson, D. H., Prater, M., & Dyches, T. T. (2013). Video self-modeling on an iPad to teach functional math skills to adolescents with autism and intellectual disability. *Focus on Autism and Other Developmental Disabilities, 28,* 67-77. doi: 10.1177/1088357613478829

Busse, R. T., McGill, R. J., & Kennedy, K. S. (2014). Methods for assessing single-case school-based intervention outcomes. *Contemporary School Psychology*. Advance online publication. doi: 10.1007/s40688-014-0025-7

Clark, E., Beck, D., Sloane, H., Goldsmith, D., Jenson, W., Bowen, J., & Kehle, T. (1993). Self-modeling with preschoolers: Is it different? *School Psychology International, 14,* 83-89. doi: 10.1177/0143034393141006

Cohen, J. S. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Collier-Meek, M. A., Fallon, L. M., Johnson, A. H., Sanetti, L. M. H., & Delcampo, M. A. (2011). Constructing self-modeling videos: Procedures and technology. *Psychology in the Schools, 49,* 3-14. doi: 10.1002/pits.20614

Cumming, G., & Finch, S. (2001). A primer on understanding, use, and calculation of confidence intervals that are based on central and noncentral distributions. *Educational and Psychological Measurement, 61,*530-572. doi: 10.1177/0013164401614002

Davis, R. A. (1979). The impact of self-modeling on problem behaviors in school-age children. *School Psychology Review, 8,* 128-132. Retrieved from https://www.nasponline.org

Dodge, K. A., & Godwin, J. (2013). Social-information processing patterns mediate the impact of preventive intervention on adolescent antisocial behavior. *Psychological Science, 24,* 456-465. doi. 10.1177/0956797612457394

Dishion, T. J., Piehler, T. F., & Myers, M. W. (2008). Dynamics and ecology of adolescent peer influence. In M. J. Prinstein & K. A. Dodge (Eds.), *Understanding peer influence in children and adolescents* (pp. 72-93). New York: Guilford Press.

Durlak, J. A. (1999). How to select, calculate, and interpret effect sizes. *Journal of Pediatric Psychology, 34,* 917-928. doi: 10.1093/jpepsy/jsp004

Foster, S. L., & Mash, E. J. (1999). Assessing social validity in clinical treatment research: Issues and procedures. *Journal of Consulting and Clinical Psychology, 67,* 308-319. doi: 10.1037/0022-006X.67.3.308

Gelbar, N. W., Anderson, C., McCarthy, S., & Buggey, T. (2012). Video self-modeling as an intervention strategy for individuals with autism spectrum disorders. *Psychology in the Schools, 49,* 15-22. doi: 10.1002/pits.20628

Hageman, W. J., & Arrindell, W. A. (1999). Establishing clinically significant change: Increment of precision and the distinction between individual and group level of analysis. *Behaviour Research and Therapy, 37,* 1169-1193.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. New York: Academic Press.

Hitchcock, C. H., Dowrick, P. W., & Prater, M. A. (2003). Video self-modeling intervention in school-based settings: A review. *Remedial and Special Education, 24,* 36-45. doi: 10.1177/074193250302400104

Hsu, L. M. (1989). Reliable changes in psychotherapy: Taking into account regression toward the mean. *Behavioral Assessment, 11,* 459-467.

Hsu, L. M. (1999). Caveats concerning comparisons of change rates obtained with five methods of identifying significant client changes. *Journal of Consulting and Clinical Psychology, 67,* 594-598. doi: 10.1037/0022-006X.67.4.594

Jacobson, N. S., Follette, W. C., & Revenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy, 15,* 336-352. doi: 10.1016/S0005-7894(84)80002-7

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

Kahn, J. S., Kehle, T. J., Jenson, W. R., & Clark, E. (1990). Comparison of cognitive-behavioral, relaxation, and self-modeling interventions for depression among middle-school students. *School Psychology Review, 19,* 196-211. Retrieved from http://www.nasponline.org

Kehle, T. J., Clark, E., Jenson, W. R., & Wampold, B. E. (1986). Effectiveness of self-observation with behavior disordered elementary school children. *School Psychology Review, 15,* 289-295. Retrieved from http://www.nasponline.org

Kehle, T. J., Madaus, M. R., Baratta, V. S., & Bray, M. A. (1998). Augmented self-modeling as a treatment for children with selective mutism. *Journal of School Psychology, 36,* 247-260. doi: 10.1016/S0022-4405(98)00013-2

Lambert, M. J., Hansen, N. B., & Bauer, S. (2007). Assessing the clinical significance of outcomes results. In A. M. Nezu & C. M. Nezu (Eds.), *Evidence-based outcomes*

*research: A practical guide to conducting randomized controlled trials for psychosocial interventions*. New York: Oxford University Press.

Legendre, P. (2005). Species associations: The Kendell coefficient of concordance revisited. *Journal of Agriculture, Biological, and Environmental Statistics, 2,* 226-245. doi: 10.1198/108571105X46642

Lunnen, K. M., & Ogles, B. A. (1998). A multiperspective, multivariable evaluation of reliable change. *Journal of Consulting and Clinical Psychology, 66,* 400-410. doi: 10.1037/0022-006X.66.2.400

Madaus, M. R., & Ruberto, L. M. (2012). Application of self-modeling to externalizing and internalizing disorders. *Psychology in the Schools, 49,* 42-51. doi: 10.1002/pits.20619

Marsden, J., Eastwood, B., Wright, C., Bradbury, C., Knight, J., & Hammond, P. (2011). How best to measure change in evaluations of treatment for substance use disorder. *Addiction, 106,* 294-302. doi: 10.1111/j.1360-0443.2010.03143.x

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

McCurdy, B. L., & Shapiro, E. S. (1988). Self-observation and the reduction of inappropriate classroom behavior. *Journal of School Psychology, 36,* 247-260. doi: 10.1016/0022-4405(88)90044-1

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7,* 19-40. Doi: 10.1037//1082-989X.7.1.19

McGill, R. J., & Busse, R. T. (2014). An evaluation of multiple single-case outcome indicators

using convergent evidence scaling. *Contemporary School Psychology, 18,* 13-23. doi: 10.1007/s40688-013-0004-4

McGlinchey, J. B., Atkins, D. C., & Jacobson, N. S. (2002). Clinical significance methods: Which one to use and how useful are they? *Behavior Therapy, 33,* 529-550. doi: 10.1016/S0005-7894(02)80015-6

Nunnally, J., & Kotsche, W. (1983). Studies of individual subjects: Logic and methods of analysis. *British Journal of Clinical Psychology, 22,* 83-93. doi: 10.1111/j.2044-8260.1983.tb00582.x

Ohtake, Y., Takeuchi, A., & Watanabe, K. (2014). Effects of video self-modeling on eliminating public undressing by elementary-aged students with developmental disabilities during urination. *Education and Training in Autism and Developmental Disabilities, 49,* 32-44. Retrieved from http://daddcec.org

Possell, L. E., Kehle, T. J., McLoughlin, C. S., & Bray, M. A. (1999). Self-modeling as an intervention to reduce disruptive classroom behavior. *Cognitive and Behavioral Practice, 6,* 99-105. doi: 10.1016/S1077-7229(99)80017-0

Reynolds, C. R., & Kamphaus, R. W. (2004a). *Behavior Assessment System for Children* (2nd ed.). San Antonio, TX: Pearson.

Reynolds, C. R., & Kamphaus, R. W. (2004b). *Behavior Assessment System for Children manual* (2nd ed.). San Antonio, TX: Pearson.

Rudy, N. A., Betz, A. M., Malone, E., Henry, J. E., & Chong, I. M. (2014). Effects of video modeling on teaching bids for joint attention to children with autism. *Behavioral Interventions, 29,* 269-285. doi: 10.1002/bin.1398

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,* 8-14. Retrieved from https://www.txasp.org

Schunk, D. H., & Hanson, A. R. (1989). Self-modeling and children's cognitive skill learning. *Journal of Educational Psychology, 81,* 155-163. doi: 10.1037/0022-0663.81.2.155

Shear, S. M., & Shapiro, E. S. (1993). Effects of using self-recording and self-observation in reducing disruptive behavior. *Journal of School Psychology, 31,* 519-534. doi: 10.1016/0022-4405(93)90035-H

Speer, D. C. (1992). Clinically significant change: Jacobson & Truax (1991) revisited. *Journal of Consulting and Clinical Psychology, 60,* 402-408. doi: 10.1037/0022-006X.60.3.402

Stein, S, & Watson, T. S. (2007). Review of the Behavior Assessment System for Children-Second Edition. In K. F. Geisinger, R. A. Spies, J. F. Carlson, & B. S. Plake (Eds.), The seventeenth mental measurements yearbook (pp. 96-102). Lincoln, NE: Buros Institute of Mental Measurements.

Thompson, B. (1999). Statistical significance tests, effect size reporting and the vain pursuit of pseudo-objectivity. *Theory & Psychology, 9,* 191-196. doi: 10.1177/095935439992007

Volker, M. A. (2006). Reporting effect size estimates in school psychology research. *Psychology in the Schools, 43,* 653-672. doi: 10.1002/pits.20176

Wilson, K. P. (2013). Incorporating video modeling into a school-based intervention for students with autism spectrum disorders. *Language, Speech, and Hearing Services in the Schools, 44,* 105-117. doi: 10.1044/0161-1461(2012/11-0098)

Zahra, D., & Hedge, C. (2010). The reliable change index: Why isn't it more popular in academic psychology? *PsyPAG Quarterly, 76,* 14-19. Retrieved from http://www.psypag.co.uk/the-quarterly/

Table 1

*Demographic Information for Intervention Participants (N =4)*

| Student | Gender | Age | Grade | Ethnicity | Eligibility | Problem Behavior(s) |
|---|---|---|---|---|---|---|
| John | Male | 8 | 2 | Caucasian | OHI<br>ED | Aggression |
| Steven | Male | 8 | 2 | Caucasian | ED<br>SI | Aggression |
| Jose | Male | 7 | 1 | Hispanic | OHI | Hyperactivity<br>Impulsivity |
| Kathy | Female | 7 | 1 | Caucasian | TBI<br>SI | Aggression<br>Following Directions |

*Note*. Student names are pseudonyms to protect confidentiality. ED = Emotional Disturbance; OHI = Other Health Impairment; SI = Speech Impairment; TBI = Traumatic Brain Injury.

Table 2

*Formulas for the Calculation of Individual Change*

| Method | Formula | |
|---|---|---|
| Jacobson-Truax Method | $\text{RCI} = \frac{X_{POST} - X_{PRE}}{\sqrt{2S_E{}^2}}$ | (1) |
| | $S_E = SD\sqrt{1 - r}$ | (2) |
| Gullikson-Lord-Novik Method | $\text{RCI} = \frac{(X_{POST} - M) - r_{xx}(X_{PRE} - M)}{SD_{PRE}\sqrt{1 - r_{xx}{}^2}}$ | (3) |
| Means Difference Effect Size | $\text{g} = \frac{\text{x}_1 - \text{x}_2}{S^*}$ | (4) |
| | $S^* = \sqrt{\frac{(n_1 - 1)s_1{}^2 + (n_2 - 1)s_2{}^2}{n_1 + n_2 - 2}}$ | (5) |

*Note*. $X_{POST}$= posttreatment score; $X_{PRE}$ = pretreatment score; $S_E$ = standard error of the measurement; *SD* = standard deviation of outcome measure; *r* = internal consistency of outcome measure; $r_{xx}$ = test-retest reliability of outcome measure; *M* = hypothesized population mean; $SD_{PRE}$ = standard deviation of pretreatment scores. $\text{x}_1$ = posttreatment mean or individual score; $\text{x}_2$ = pretreatment mean or individual score; $S^*$ = pooled standard deviation; $n_1$ = pretreatment participants; $n_2$ = posttreatment participants; $s_1$ = standard deviation of pretreatment scores; $s_2$ = standard deviation of posttreatment scores.

Table 3

*Descriptive Statistics for Behavior Assessment System for Children-Second Edition (BASC-2) Externalizing Problems Composite Outcomes*

| | Teacher Rating Scales (TRS) | | | Parent Rating Scales (PRS) | | |
|---|---|---|---|---|---|---|
| Student | T-Score $_{PRE}$ | T-Score $_{POST}$ | *D* | T-Score $_{PRE}$ | T-Score $_{POST}$ | *D* |
| John | 79 | 67 | -12 | 76 | 54 | -22 |
| Steven | 88 | 77 | -11 | 62 | 62 | 0 |
| Jose | 72 | 62 | -10 | 53 | 56 | +3 |
| Kathy | 70 | 65 | -5 | 64 | 62 | -2 |
| **Mean** | *77.25* | 67.75 | -9.50 | *63.75* | 58.50 | -5.25 |
| ***SD*** | 8.14 | 6.50 | 3.11 | 9.46 | 4.12 | 11.35 |

*Note*. *D* = difference score. *Italics* indicate pretreatment mean values utilized to determine cutoff points for the clinical significance criteria. Hedges' g (TRS) = 1.12**, 95% CI [-3.98, 6.23]. Hedges' g (PRS) = 0.63*, 95% CI [-4.43, 5.68].

* Moderate effect size. ** Large effect size.

Table 4

*Individual Outcomes across Multiple Statistical and Clinical Change Indices*

| | RCI | | MDES | | | Cutoff Criteria | |
|---|---|---|---|---|---|---|---|
| TRS | JT | GLN | *g* | 95% CI | Clinical | Nominal | Status |
| John | **-4.90** | **-2.56** | 1.42** | [-3.69, 6.52] | No | No | I |
| Steven | **-4.49** | **-2.03** | 1.39** | [-3.81, 6.40] | No | No | I |
| Jose | **-4.08** | **-2.20** | 1.18** | [-3.92, 6.23] | No | No | I |
| Kathy | **-2.04** | -0.85 | 0.59* | [-4.51, 5.69] | No | No | I |

| | RCI | | MDES | | | Cutoff Criteria | |
|---|---|---|---|---|---|---|---|
| PRS | JT | GLN | g | 95% CI | Clinical | Nominal | Status |
| John | **-5.88** | **-4.97** | 2.62** | [-2.43, 7.68] | No | Yes | R |
| Steven | 0.00 | 0.31 | 0.00 | [-5.06, 5.06] | No | No | U |
| Jose | 0.80 | 0.93 | -0.34 | [-5.41, 4.70] | No | Yes | D |
| Kathy | -0.53 | -0.15 | 0.24 | [-4.82, 5.29] | No | No | U |

*Note*. TRS = teacher rating scale; PRS = parent rating scale; RCI = reliable change index; MDES = means difference effect size; CI = confidence interval. JT = Jacobson-Truax Method; GLN = Gullikson-Lord-Novick Method. **Bold** denotes RCI values that exceed thresholds for reliable change (e.g., $p \leq .05$). R = recovered; I = improved; U = unchanged; D = deteriorated.

*Moderate Effect Size. **Large Effect Size

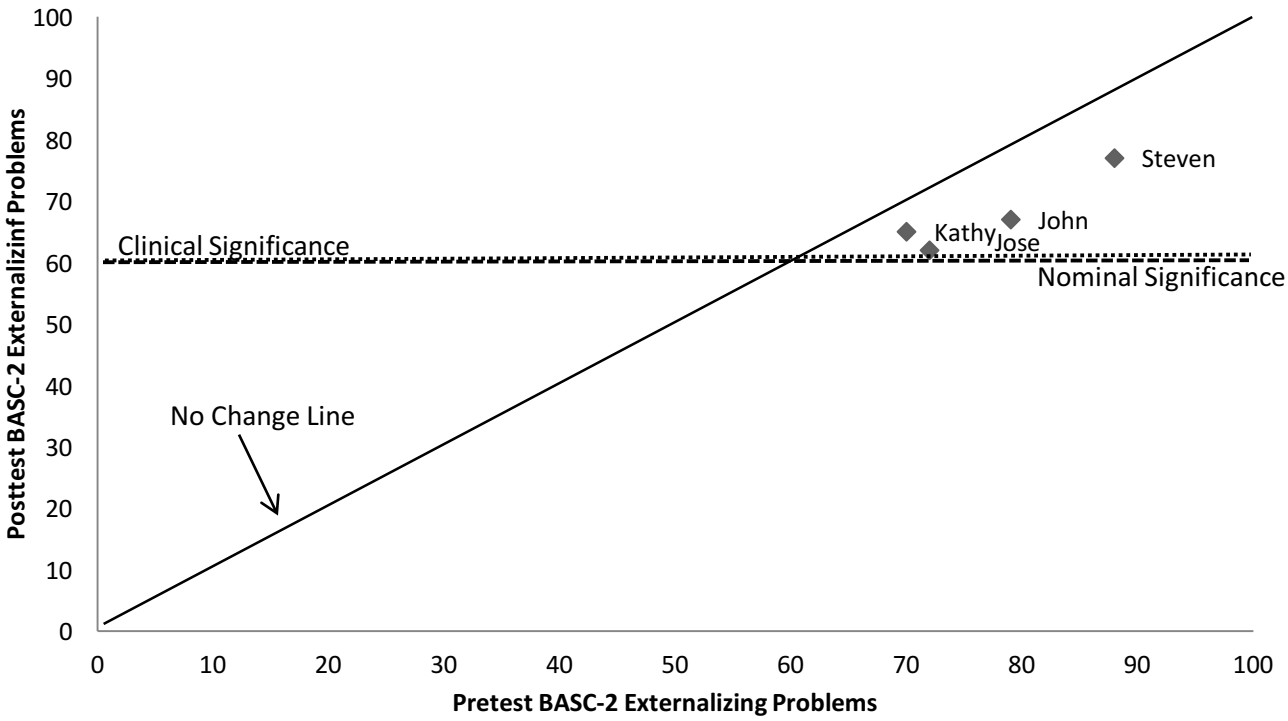


*Figure 1*. BASC-2 Teacher Rating Scales (TRS) scores showing measurement variation bounds and cutoff points for multiple levels of clinical significance. Nominal significance corresponds to posttest scores at or below 59. Clinical significance corresponds to scores at or below 61 (e.g., Jacobson & Truax, 1991).

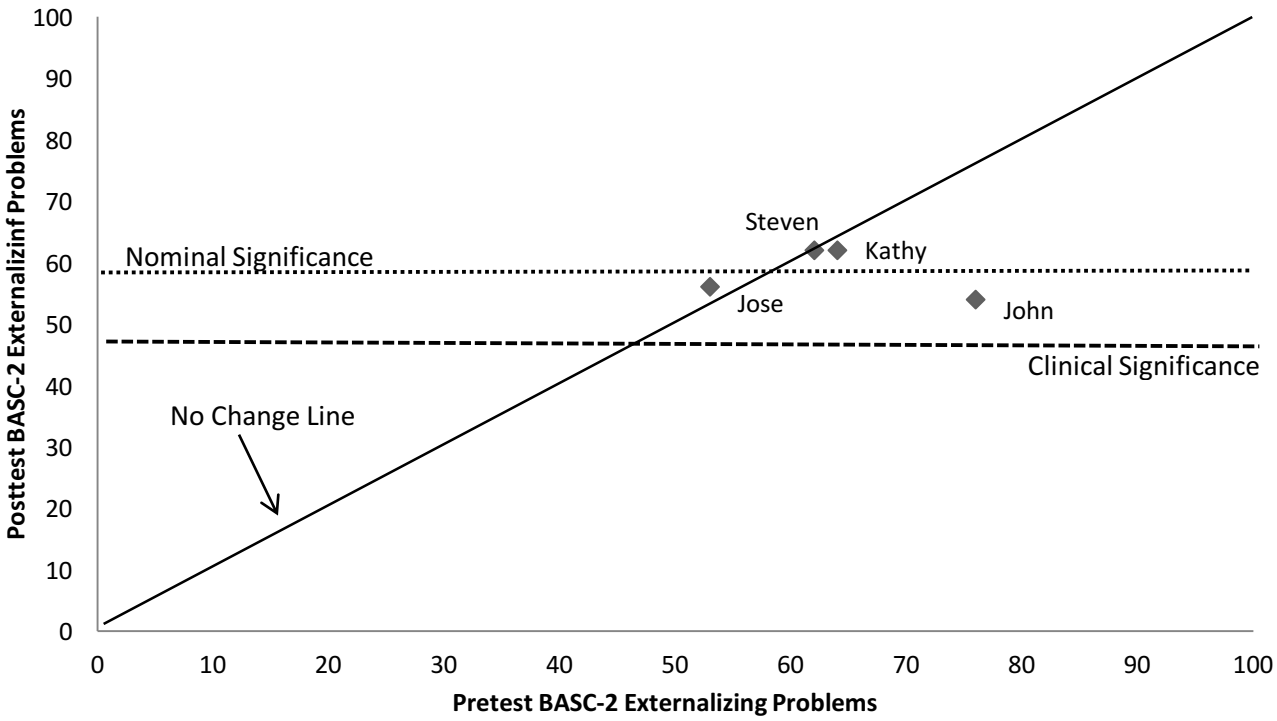


*Figure 2*. BASC-2 Parent Rating Scales (PRS) scores showing measurement variation bounds and cutoff points for multiple levels of clinical significance. Nominal significance corresponds to posttest scores at or below 59. Clinical significance corresponds to scores at or below 45 (e.g., Jacobson & Truax, 1991).