Utah State University

DigitalCommons@USU

All Graduate Theses and Dissertations | Graduate Studies

8-2012

Agreement of Peer Comparison Data Between Direct Behavior Rating Scales and Systematic Direct Observation Methods

Elizabeth Ashley Popescue
Utah State University

Follow this and additional works at: https://digitalcommons.usu.edu/etd

 Part of the Psychology Commons

Recommended Citation

Popescue, Elizabeth Ashley, "Agreement of Peer Comparison Data Between Direct Behavior Rating Scales and Systematic Direct Observation Methods" (2012). *All Graduate Theses and Dissertations*. 1305.
https://digitalcommons.usu.edu/etd/1305

This Thesis is brought to you for free and open access by the Graduate Studies at DigitalCommons@USU. It has been accepted for inclusion in All Graduate Theses and Dissertations by an authorized administrator of DigitalCommons@USU. For more information, please contact digitalcommons@usu.edu.

AGREEMENT OF PEER COMPARISON DATA BETWEEN DIRECT BEHAVIOR RATING SCALES AND SYSTEMATIC DIRECT OBSERVATION METHODS

by

Elizabeth Ashley Popescue

A thesis submitted in partial fulfillment
of the requirements for the degree

of

EDUCATIONAL SPECIALIST

in

Psychology

Approved:

_____	_____
Donna Gilbertson, Ph.D. Major Professor	Gretchen Gimpel Peacock, Ph.D. Committee Member
_____	_____
Kyle Hancock, Ph.D. Committee Member	Mark. R. McLellan, Ph.D. Vice President for Research and Dean of the School of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2012

Copyright © Elizabeth Popescue 2012

All Rights Reserved

ABSTRACT

Agreement of Peer Comparison Data Between Direct Behavior Rating Scales and Systematic Direct Observation Methods

by

Elizabeth Ashley Popescue, Educational Specialist

Utah State University, 2012

Major Professor: Donna Gilbertson, Ph.D.
Department: Psychology

Recently, Direct Behavior Ratings have been shown to be a promising new tool for observing students and classrooms in an education setting for a variety of behaviors. The traditional method of observing students and classroom behavior was through tools called Systemic Direct Observations. Currently, there are only a few studies looking at the use of a Direct Behavior Rating as a device to collect peer comparison data to estimate classwide behavior problems. This study examined the estimated percentages of on-task and disruptive behavior between a Systemic Direct Observation with momentary time sampling and three random peers, a Systemic Direct Observation with momentary time sampling using the entire class, and a Direct Behavior Rating. Multiple undergraduate classrooms were taped and divided up into twenty-five 7-minute segments. The videos were then coded on all three of the observation forms with 100% reliability ratings. Results indicated that there was a strong relationship between the Direct Behavior Rating and the SDO classwide on-task estimates with 37% of the variance in

the Systemic Direct Observation classwide data consistent with the Direct Behavior Rating data. There was a moderate relationship between the on-task Direct Behavior Rating and three-peer on-task with 13% of the variance in the Systemic Direct Observation data as a portion of the Direct Behavior Rating data. Results also showed that there was a significant correlation between Direct Behavior Rating both of the Systemic Direct Observation methods with 43% for the classwide Systemic Direct Observation and 39% of the three-peers Systemic Direct Observation variance consistent with the Direct Behavior Rating data. Implications and future directions were considered. The research yielded results that indicated that Direct Behavior Ratings might be a useful tool when evaluating classwide behavior, and that further research is warranted.

(86 pages)

PUBLIC ABSTRACT

Agreement of Peer Comparison Data Between Direct Behavior Rating Scales and Systematic Direct Observation Methods

by

Elizabeth Ashley Popescue

Utah State University, 2012

Elizabeth Popescue, M.A. and Donna Gilbertson, Ph.D. at Utah State University (USU) propose to evaluate how a new type of tool to measure classwide behavior in a school classroom compares to the current gold-standard tool. They will coordinate with undergraduate professors teaching psychology classes to be able to videotape classrooms of students to use as subjects to observe behavior with the different types of tools. The advantage of the new tool is the reduced amount of effort and time it takes to collect the behaviors observed. This could lead to a more efficient way for school personnel to collect data needed from classrooms to make decisions if it compares well with current tools. This project team proposes a one-year project with no anticipated costs.

ACKNOWLEDGMENTS

This thesis would not have been possible without the support of many people. I would like to thank my thesis advisor, Dr. Donna Gilbertson, for her constant support, guidance, and insight throughout this entire process. I would also like to acknowledge my committee members, Dr. Gretchen Peacock and Dr. Kyle Hancock, for providing their assistance, helpful suggestions, and time.

I would also like to thank Ryan Greene, Joanna Jenkins, and Evan Adams for giving up their valuable time to set up equipment and sit through classes to tape and to help me and code videos during late nights and early mornings. Also, thank you to Allison White for opening up her classes to me to tape. Special thanks to my family, friends, and colleagues for encouraging me through this process and having extra confidence for me when I needed it.

Elizabeth Ashley Popescue

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Direct observations of student behavior in the classroom setting are one of the most common behavior assessment tools employed by school psychologists to identify and analyze problem behaviors in the classroom (Chafouleas, Riley-Tillman, & Christ, 2009). Most often, a concern about a possible behavior problem with a child will spur a classwide direct observation (Eidle, Truscott, Meyers, & Boyd, 1998). Estimates of the prevalence of students presenting behavior problems fall between 2% and 16% (Kauffman, 2001). Behavioral problems are a common issue that interferes with learning with estimates of 50% of teacher referrals for intervention services being due to behavior problems (Bramlett, Murphy, Johnson, Wallingsford, & Hall, 2002; Eidle et al., 1998). An important purpose of observational data collected when a child is initially referred due to behavior problems is to identify whether or not the child's behavior is atypical of peers and to confirm that effective classroom management is in place. This is accomplished by collecting observational data to estimate the behavior of peers who are provided with the same classroom instruction and behavior management strategies. Comparison peers are used to determine when the target student's behavior is atypical relative to other students' behavior in the class and to determine when classroom management is suitable to maintain appropriate classroom behavior (DuPaul & Hoff, 1998; Salvia & Hughes, 1990). If a classwide problem is not ruled out first, then those children that are being specifically singled out for observation may appear to be more of a problem than they really are (Deno, 1980). This distinction between an individual

versus classwide behavior problem is important when developing a hypothesis of the function of a child's behavior problem to select the type of effective intervention needed to resolve the problem. In the case of a classwide behavior problem with many students exhibiting problems in a referred child's classroom, a classwide intervention that employs effective management may effectively reduce disruptive behaviors for the referred child as well as peers. In well-managed classrooms with observed individual behavior problems, an individual intervention is needed to reduce disruptive behaviors (Ellis & Magee, 2004).

There are a variety of techniques to collect data from comparison peers. Time sampling techniques such as partial, whole, or momentary are common practices to collect peer comparison data (Volpe, DiPerna, Hintze, & Shapiro, 2005). Each of these methods are direct observation measures employed by school psychologists, commonly called Systematic Direct Observations (SDO). Each method has its own strengths and weaknesses, but there is no standard for which of these should be used (Bell & Barnett, 1999). The manner that the peer is chosen, how many peers are chosen, the sex of the peer, and the order in which the peers are chosen also varies greatly. Yet, knowledge on how accurate each of these SDO techniques is in determining normative data is a critical factor when making important behavioral and educational decisions for a student. Moreover, the SDO employed by school psychologists are complex, time consuming, and take the full attention of the school psychologist, only allowing him/her to complete an observation on one student at a time. Given the resources required to conduct SDOs, a group of researchers has recently examined the utility of Direct Behavior Ratings (DBR)

as a simpler, alternative method to observe classroom behavior through multiple different studies. This group of researchers conducted multiple studies showing that DBRs are a reliable supportive measure for assessing classwide, as well as individual student behaviors in multiple areas including on-task behavior, disruptive behavior, and engagement. Other findings included that training is an important key when conducting DBRs, that anchors for rating behaviors do not affect the reliability of the scores and that they are a valid way of measuring behaviors in a classroom (Chafouleas et al., 2010; Chafouleas, McDougal, Riley-Tillman, Panahon, & Hilt, 2005; Riley-Tillman, Chafouleas, Sassu, Chanese, & Glazer, 2008; Riley-Tillman, Methe, & Weegar, 2009); Schlientz, Riley-Tillman, Briesch, Walcott, & Chafouleas, 2009). DBR is composed of a likert-type rating scales on a few specific behaviors that can be used to gather information about the level of or change in behaviors for either individual students or the class as a whole. Research on the correspondence of DBR ratings with SDO estimates when observing individual behaviors suggests potential for providing information about the class level to determine how much the class is exhibiting disruptive behavior to screen out the need for further, more intensive, individualized assessment when it is not actually needed.

Given the number of students experiencing behavior difficulties who disrupt learning (Eidle et al., 1998) and the variability in reasons for these difficulties, an important area of research is to examine the validity and reliability of observation assessment methods that can be easily used in classroom settings to quickly gather information about reasons why behavior problems are occurring to develop the

interventions that would best remediate the problem. Although several types of observation methods can be used to collect peer comparison, preliminary support suggests that DBR is a promising, simple observation method to gather useful information about student behavior (Chafouleas et al., 2005, 2010; Riley-Tillman et al. 2008; Riley-Tillman, Methe et al., 2009; Schlientz, Riley-Tillman, Briesch, Walcott, & Chafouleas, 2009) with preliminary support for assessing classwide behavior to determine whether or not a classwide intervention is needed (Riley-Tillman, Chafouleas, Christ, Briesch, & LeBel, 2009). However, additional research is needed to determine the relation of DBRs compared to momentary time sampling procedures (given its higher degree of accuracy in estimating individual behavior relative to other interval recording systems) to make decisions about behavior management at the class level (Hintze, 2004; Salvia & Huges, 1990). Thus, the purpose of the present study will be to expand the concurrent validation of DBRs conducted by Chafouleas et al. (2005) and Riley-Tillman et al. (2008), by investigating concurrent validation of DBRs for estimating classwide behavior. The present study is designed to assess the agreement between behavior data collected using two types of peer comparison Systematic Direct Observation (SDO) procedures and DBRs.

CHAPTER II

REVIEW OF LITERATURE

Prevalence of Behavior Problems in Classroom Setting

Children across the United States suffer from a variety of mental health problems. It has been estimated that as many as one-third of all children are not learning to their full potential in school because they have some type of psychosocial problem that interferes with the learning process that requires an intervention (Epstein, Atkins, Cullinan, Kutash, & Weaver, 2008). These problems are often categorized into two different classes: externalizing problems and internalizing problems. Externalizing problems concern behaviors that the child outwardly displays, that are often disruptive, while internalizing problems are inward-felt affective states that are less likely to be noticed by outsiders (Gimpel & Holland, 2003).

In schools, externalizing behavior problems are a major concern since these behaviors interrupt students' academic and social performance (Epstein et al., 2008). While internalizing problems can cause these issues as well, but externalizing problems are often more of a concern since they are outwardly disruptive and can affect other students in the classroom as well. Behavior problems and disorders are considered "externalizing problems" because evidence of these problems is often observable, as well as disruptive, defiant, and sometimes dangerous (Merrell, 2007). External behavior problems can range from aggression, noncompliance, and disruptive behavior such as talking out of turn or off-task activity, to more severe disorders such as oppositional

defiant disorder and attention deficit/hyperactivity disorder.

While behavior problems continue to be an extensive problem in schools, actual prevalence numbers are difficult to uncover because there is controversy within the field about whether behavior problems should even be diagnosed in children, and if so, what the criteria should be to make the diagnosis (Gimpel & Holland, 2003). One estimate has put the prevalence behavior problems in all children between 2% and 16% (Kauffman, 2001). Others put the figure for emotional and behavior disorders in school-aged children at 20% (Satcher, 1999). On the younger end of the spectrum, it has been shown that of all 2 and 3 year olds, around 25% may display some type of severe behavior problem and of those, nearly 50% will fail to grow out of it and continue to display the behavior problems (Campbell, 1995; Lavigne et al., 1998). Furthermore, it has been estimated, based on teacher ratings that for 3-5 year olds, the prevalence of behavior problems is between 14% and 52% (Qi & Kaiser, 2003). While most children fall at the lower end of the behavior problem spectrum, it has been estimated that among primary and secondary students in the United States diagnosed with a behavior disorder, around 5% to 16% will fall in the severe behavior problem category (Bowen, Jensen, & Clark, 2004).

Many children will show some type of behavior problem at one time or another, but it is those children that chronically and consistently fall on the severe side of the spectrum and disrupt learning that are of most concern in the school environment (Bowen et al., 2004). There are some data that show that behavior problems should be of great concern to parents and teachers because their onset can occur very early and they can be

stable and persistent through life for about 50% of children when observed at a young age (Gimpel & Holland, 2003).

Behavior problems are a primary reason why a teacher will refer a child to be observed in order to identify the type of problem and intervention needed (DuPaul & Stoner, 2003). Results of a survey study conducted by Bramlett et al. (2002) found that school-based team members reported that 39% of referrals to school-based problem solving teams are related to task-completion, 26% are related to conduct issues, 24% are related to motivation, 17% are related to defiance, and 6% are related to violence. Those that have these severe behavior problems may have a diagnosable behavior disorder. According to the *Diagnostic and Statistical Manual of Mental Disorders-IV-TR,* it is estimated that 3%-7% of children in schools may have one of the three types of attention-deficit/hyperactivity disorder, 1-10% may have conduct disorder, and 2-16% may have oppositional defiant disorder (American Psychiatric Association, 2000). Boys are also more likely to display these behavior disorders than girls, and when diagnosed early, they are also more likely to have the problem behavior be persistent (Campbell, 1995).

The high prevalence of behavior problems in schools often has multiple repercussions. Those children that display behavior problems have been shown to have long-term negative consequences, such as greater difficulty with academics and social interactions with other peers as well as teachers (Evertson & Weinstein, 2003). Behavior problems cause many different issues in a classroom and it is one of the most likely reasons a teacher would refer a child to the school psychologist for observation (DuPaul & Stone, 2003). Often, children with behavior problems experience school as an

unpleasant place to be (Bowen et al., 2004). Early identification is key for students with possible behavior problems so that adequate support can be made available to them. Having a combination of behavior problems and academic problems or social problems leads to a dangerous road of possible school avoidance, failure or dropout (Bowen et al., 2004). To prevent these issues, students with behavior problems need to be identified early so support can be implemented.

Classroom Management Issues

Students are often observed in the classroom to identify environmental factors that are related to the occurrence of inappropriate behaviors. Several reviews of effective intervention for individual behavior problems suggest that there are various common reasons for the causes of behavior problems in a classroom setting (Gresham, Watson, & Skinner, 2001). A poorly managed classroom is often a primary reason why a student misbehaves (Huston-Stein, Friedrich-Cofer, & Susman, 1977; Simonsen, Fairbanks, Briesch, Myers, & Sugai, 2008; Susman, Huston-Stein, & Friedrich-Coffer, 1980). Often, a classroom that is not being managed properly will result in undesired behaviors from many children (Hieneman, Dunlap, & Kincaid, 2005). If an individual rather than a classwide intervention is not in place, then many students may continue to struggle and show issues with behavior. Multiple children in a poorly managed classroom will exhibit disruptive behavior problems making it difficult to identify children who would benefit from a classwide behavior management intervention or those that require a more individualized intensive intervention. Thus, observation assessment to identify reasons

for an individual student's behavior problems should be conducted on a classwide level as well as the individual level to first identify or rule out classwide management as a plausible reason for individual behavior problems. In the case of a classwide behavior problem with many students exhibiting problems in a referred child's classroom, there are many studies suggesting effective strategies that can be implemented for a classwide behavior management problem to increase appropriate behavior for the majority of the students in the class (Conroy, Stichter, Daunic, & Haydon, 2008; Handler et al., 2007).

Behavioral Observation Practice and Purpose

Classroom teachers seek the help of an education professional, such as a school psychologist, when they are having behavior problems with students (Shapiro & Clemens, 2007). In schools, assessments of referred behavior problems frequently include a classroom observation (Crone & Horner, 2003). Wilson and Reschly (1996) surveyed over 1,000 school psychologists throughout the United States and found that the assessment most administered by school psychologists was a structured classroom observation. On average, the surveyed school psychologists completed 15 observations a month. In a later survey of 648 nationwide school psychologists conducted by Shapiro and Heick (2004), the reported use of behavior assessments increased and nearly half (47.8%) claimed that they used a direct behavior observation at least 8-10 times out of their last 10 cases. Classroom observations are typically conducted for 15-20 minutes and should be as unobtrusive as possible with the observer sitting in the rear of the classroom (Chafouleas, Riley-Tillman, & Sugai, 2007; Forness & Esveldt, 1975a).

Behavior observations are an important part of behavior assessments because they are an objective, empirically validated measure to gather information about behavior problems (Merrell, 2007). Direct observations offer a unique in-vivo look at the child in his or her natural environment, which can often yield beneficial data. For example, data can be used to identify the type of behavior problem, the severity of the behavior problem, and the reasons why behavior problems are occurring to determine intervention needs. Identified environmental variables maintaining problem behavior that can be altered, eliminated, or reversed to reduce problem behaviors and increase appropriate behaviors (Landau & Swerdlik, 2005).

Classroom observations may also be a useful assessment method to determine if the problem requires an individual or classwide behavior management intervention (Chafouleas et al., 2007). One common observation method often used to determine if the referred problem is an individualized or classwide behavior problem is to compare a child's behavior with a peer comparison in the classroom. In doing so, a school psychologist can determine if the peers are well managed and if severity of the problem being displayed by the target student is within a more severe range of what is being displayed by other students in the classroom (Crone & Horner, 2003). These data also provide a baseline behavior performance to compare the effects of future interventions on expected behavior change in the classroom environment (Miltenberger, 2005).

Salvia, Ysseldyke, and Bolt (2009) suggested that all classroom observation methods meet five goals: measure specific behaviors, operationally define the behaviors, select the time and place of the observation with a rationale, use objective procedures to

collect data, and score and summarize data collected using standardized rules. In addition to these five goals, Miltenberger (2005) has suggested that choosing the correct type of recording procedure and the correct type of instrument in which to measure the behaviors are critical factors to collecting useful data for accurate decision-making. There are various systematic direct observation methods that can be selected and modified to best meet the proposed observation goals. A brief discussion of structured or systematic observational methods that are most frequently used by school psychologist is included in the following section.

Systematic Behavior Observation Recording Systems

Several observation methods can be used to gain information about external behaviors in the context in which they occur. For example, event recording can be used to estimate the frequency of behavior occurrence by accounting or tallying the number of times a behavior occurs during a specific amount of time. Duration recording can be used to determine how long a behavior continuously occurs by recording the amount of time a behavior occurs during an observation session. Interval recording is an observation method that can be used to estimate both the number of times and the duration that a behavior occurs during a specified time interval (Cooper, Heron, & Heward, 2007). Interval recording is the measure of the absence or presence of a behavior during brief, but equal time intervals. For example, the presence or absence of a behavior may be recorded every 10 seconds for a 15-minute observation session such that the behavior presence or absence is recorded a total of 150 times. Behavior occurrence is then

estimated as the percent of time that the behavior was present during the observation. Furthermore, interval-recording can be broken down into two subtypes: partial-interval recording, where a behavior is recorded if it occurs anytime during the interval and whole-interval recording, where a behavior is only recorded if it occurs during the whole interval. Finally, momentary time sampling is similar to interval recording, except that a behavior is recorded as present or absent only if it occurs at the end of a specified time interval (Salvia et al., 2009).

While all of these methods are currently used in schools, each method has specific limitations (Cooper et al., 2007). For example, while event recording is effective at recording the frequency of the behavior, it lacks the ability to record the durations. It is also best utilized with behaviors that are discrete (Hintze, 2004). With duration recording, temporal features of the behavior occurring can be recorded, but it is best used with behaviors that have long durations, as it can be difficult to continually time behaviors that have a short duration (Shapiro & Clemens, 2007).

Many school-based assessments are conducted using an interval recording system because this observation method provides a more accurate estimate of not only the frequency of a behavior, but the duration as well (Cooper et al., 2007). However, partial-interval recording tends to overestimate the occurrence of a behavior since the presence of a behavior is coded if it occurs in any amount of time during an interval. Alternatively, whole-interval recording tends to underestimate the occurrence of a behavior given that behavior incidences are recorded only if it is observed in an entire interval (Cooper et al., 2007; Green, McCoy, Burns, & Smith, 1982; Hintze, 2004). Furthermore, Salvia and

Hughes (1990) found that whole-interval and partial-interval were both inaccurate as far as estimating the duration and frequency. They also found that momentary-time sampling was the least biased and most accurate of all the methods used, especially when it is used in shorter intervals of 10 to 15 seconds. Green et al. (1982) compared the accuracy of whole interval, partial interval, and momentary time sampling techniques with low, intermediate, and high rates of target behavior. Fifty-four undergraduate participants collected data from a 48-minute videotape of a woman twisting her hair. The videotape was broken into 8-minute segments in which the duration of hair twisting varied. The results showed that whole and partial interval recording had poorer between-methods accuracy than did momentary time sampling. When the behavior occurred 25 times during the tape, the partial-interval method overestimated by 25%, the whole-interval method underestimated by 17%, and the momentary time sampling overestimated by 4%. When there were 50 occurrences of the behavior, the partial-interval method overestimated by 17%, the whole-interval method underestimated by 15%, and the momentary time sampling neither overestimated nor underestimated. Lastly, when there were 75 occurrences of the behavior during the tape, the partial-interval method overestimated by 10%, the whole-interval method underestimated by 23%, and the momentary time sampling underestimated by 10%. Momentary time sampling and whole-interval recording also had greater within-method accuracy than did partial-interval recording.

Despite the research supporting the use of momentary time sampling procedures to gain information about student behavior, a major obstacle of SDOs is that they can be

very time consuming, which may also drain resources and prevent a school psychologist from helping more students (Riley-Tillman et al., 2008). Often, the observer must focus solely on the one target child, even though there may be multiple children in a classroom needing observation and each observation can last anywhere from 10 to 30 minutes (Riley-Tillman et al., 2008). Furthermore, an independent observer must complete SDOs since the classroom teacher cannot teach and collect data in the systemic way that SDOs require. Having an independent observer in the room, such as a school psychologist, raises the chances that atypical behaviors will be displayed and a true estimate of typical behavior may not be obtained (Riley-Tillman et al., 2008). Given these concerns, an alternative, more efficient observation method would be useful for educators to gain useful information about student behavior.

Direct Behavior Ratings

Recently an assessment tool called a Direct Behavior Rating (DBR) has been introduced as an alternative direct SDO observation recording method to examine classroom behavior (Riley-Tillman et al., 2008). According to Chafouleas, Riley-Tillman et al. (2009), both DBRs and SDOs are methods that collect data on behaviors as it occurs in a setting that can be repeatedly recorded using standardized procedures. Basically, a DBR is a method where an observer rates some parameter of a behavior(s) (e.g., frequency, duration, percentage) that is observed in the environmental context of interest for some specified amount of time and then recorded using a rating scale format. Results of a survey distributed by Chafouleas, Riley-Tillman, and Sassu (2006) to a random

sample of 1,000 teachers across the United States showed that teachers are extremely accepting of DBRs. Of the 123 responders, 60% reported that they used a DBR in some way in their own classroom.

Given that potential advantages of DBRs relative to SDOs include simplicity, flexibility, repeatability, and time efficiency, a group of researchers have conducted a series of recent studies to investigate the reliability and validity of DBR instruments to estimate student classroom behavior. For example, Chafouleas, Christ, and Riley-Tillman (2009), found that the number of gradients used to rate a parameter of a behavior on a DBR does not substantially contribute to the variance in scores by raters, although Chafouleas, Kilgus, and Hernandez (2009) recommended that a 0-10 point gradient would be a simple measure to use for most observers. Riley-Tillman, Chafouleas et al. (2009) further compared observation ratings on DBRs by 145 undergraduate students that presented specific versus global and positive versus negative wording on the accuracy of DBRs relative to a true score. Two researchers using a computer system to record behaviors during one-second coding intervals calculated the true scores in this study. Results showed an interaction effect, $F(1,145) = 14.85$, $p = .001$, $\eta^2 = .09$, that supported that accuracy of the actual behavior compared with the true score was higher when the items on the DBR for academic engagement were defined with a global definition. There was also a significant main effect, $F(1,127) = 9.56$, $p = .002$, $\eta^2 = .07$, that supported accuracy was higher for disruptive behavior than when it was presented with a global definition and either positive or negative wording. Christ, Riley-Tillman, Chafouleas, and Jaffery (2011) conducted a study extending findings from this study by

examining positive vs. negative directionality of wording of items on a DBR rating sheet for compliance, disruption, and engagement (academically engage vs. academically unengaged, well-behaved vs. disruptive, etc.). In this study, 88 undergraduate students rated the behavior of a student in five different video clips. Each 2-minute video clip showed eight elementary school students in a classroom setting. SDO data were also collected using the Multi-Option Observation System for Experimental Studies (MOOSES) computer-based program to record behavior event using a 1-second coding intervals. Spearman rank order reliability coefficients calculated to evaluate inter-rater reliability for the DBR were .81 for positive wording and .79 for negative wording of disruptive behavior, and .75 for positive wording and .61 for negative wording of academic engagement. Correlations between the DBR with the SDO for academically engaged and disruptive behaviors ranged from 67-78. Results also showed that positive wording for on-task behavior yielded slightly less rater bias and rater error. For disruptive behavior vs. well behaved, results were mixed showing an overestimation for negative wording and underestimation for positive wording by about two points compared to the master raters who completed the SDOs. The data from the DBRs tended to create a slightly more pessimistic view of the student than did the SDO data.

Riley-Tillman, Methe, et al. (2009) examined accuracy and procedure for proportional verses absolute anchoring and observation length for DBR ratings of academically engaged behavior and disruptive behavior. For the proportional scale, 81 undergraduate students rated one target student on the estimated percentage of time the behavior was displayed, while the absolute scale required the raters to estimate the actual

amount of time a student displayed the target behaviors. Eight 5-minute video clips of a target student were obtained for this study. Observation times varied from 5 minutes to 10, 20, and 40. All participating raters rated all of the videos on both types of rating scale anchors. SDO data were also collected as the true or actual rating of behavior for each observation using one-second intervals recorded with the computer-based (MOOSES) data collection software. Results from tests of repeated measures multivariate analysis of variance (MANOVA) examining combinations of DBR rated academic engagement and disruptive behavior showed that there was a positive bias for the DBR data (ratings were overestimated by one to two points above the actual rating completed by two graduate students). There was not a significant difference between the types of anchoring systems used for any of the times (5 minutes, 10, 20, or 40). However, the duration of the observation can affect how accurate the DBR data are, but the finding varied across target behavior. These researchers found that the longer the observation session, the higher the overestimation was for disruptive behavior.

Several studies were also conducted to investigate the effect of the observer on DBR ratings. For example, Chafouleas et al. (2010) investigated the variance of the ratings across different target students, raters, and occasions of day when using a 10-point direct behavior rating form to measure academic engagement and disruptive behavior of middle school students. Data collected with seven target students, four raters, six days, three ratings, and two behaviors were compared. Two teachers and two researchers each rated one student participant at a time for ten minutes. They completed three of these observations a day for six days in a row for a total of 18 observations. Results showed

that behavior ratings between teachers and researchers did not differ substantially showing a 5% or less proportion of variance attributable to rater for both behaviors. This low proportion of variance attributable to rater was also noted with ratings of students with very high engagement or low disruptive behavior. When examining the change in generalizability and dependability coefficients as one rating was collected per day by one observer was collected over 20 observation days, results revealed that high levels of reliability (i.e., .90) were achieved at about 15 days and sufficient levels of reliability (i.e., .80) at 10 days.

Observer training effects on the accuracy of DBRs have also been investigated by Schlientz et al. (2009). In this study, 59 undergraduate students rated visually distracted and active task-manipulation behaviors of elementary school-aged children. Two groups of raters (one with training on DBRs and one without) completed 1-minute observations of these videos and rated the behaviors. Results showed that the group that received training was significantly more accurate than those that did not receive training. Chafouleas et al. (2005) conducted a study comparing data obtained between DBRs and SDOs when conducted by an external observer. Thirty-two elementary school teachers, including five special education teachers, participated in their study. Each teacher was asked to pick a student with mild behavior problems and one with a severe behavior problem, so that noncompliant behavior, disruptive behavior, and negative peer interaction could be observed. A 5-point likert-scale was used for the DBR (range 0-5) with a descriptor (occasionally) and a percentage (1-20%). Each teacher was assigned to one of four conditions (mild behavior problem with or without training or severe

behavior problem with or without training). The teachers then completed a DBR after a 15-minute observation while the outside observer completed the direct observation with 20-second intervals using momentary time sampling during the 15 minutes. A moderate degree of similarity (range, r = .48-.65) between the percentage ratings of the two behaviors on the SDO and DBR was found even when completed by separate raters. Finally, Riley-Tillman et al. (2008) conducted a similar study to examine the degree that SDOs and DBRs were correlated with each other when looking at disruptive behavior and on-task behavior. A 0-5 point likert-scale DBR and 20-second interval momentary time sampling direct observation was employed during 15-minute classroom observations. Fifteen teachers participated in this study, with 6 of them in general education classrooms and 9 in inclusive general and special education classrooms. Each teacher chose one student from his or her classroom that they believed showed disruptive behavior. Trained graduate student researchers conducted the SDO during the observation while the teacher completed the DBR after the observation period. Results revealed a moderate degree of agreement (r = .81 for on-task rating and r = .87 for disruptive behavior ratings) between the SDO completed by an outside observer and a DBR by the teacher.

Recently, Riley-Tillman, et al, (2009) investigated the degree that DBRs and SDOs observations correspond on measures of classwide engagement. In this study, data were collected as a first grade teacher implemented a classwide intervention to increase silent reading. A single case B-A-B-A reversal design was used to track the agreement of the DBR and SDO across phases. Researchers and the teacher of 14 students collected

behavior data during individual reading time using a 10-point DBR, which measured academic engagement as well as disruptive behavior. The two behaviors were rated immediately at the end of reading time. Researchers also observed behavior at the same time using a 10-second interval SDO and observed each student for one interval until all students were observed and repeated the cycle throughout the reading time. Results showed that differences between SDO and DBR phase means ranged between 2-8%. Furthermore, both measures suggested that the reading intervention successfully increased on-task behavior. Although these are promising results for the utility of DBR to assess classwide behavior, clearly more research is needed to investigate the degree of variations between the two measures when collected in other classroom settings to further assess the validity of DBR when used to collect observation data at the classwide level.

In sum, studies indicate DBRs are a viable assessment tool for observing classroom behaviors. They can be less time-consuming and more flexible, but still provide repeated data points of observation with multiple students being able to be observed at a time (Chafouleas, Riley-Tillman, et al., 2009). Only brief training is needed to complete a DBR as opposed to a SDO (Schlientz et al., 2009).

It is also important to note the limitations to using a DBR. Ratings on multiple observations should be conducted (if possible) by the person that has the most contact with the child, who is most often the teacher and not a school psychologist (Riley-Tillman et al., 2008; Riley-Tillman, Chafouleas, et al., 2009). However, many studies have demonstrated that there is good reliability when behaviors are rated by different or multiple raters (Chafouleas et al., 2010; Christ, Riley-Tillman, Chafouleas, & Boice,

2010). Furthermore, DBRs can be somewhat subjective and be affected by the halo effect, observer drift, and reactivity effects (Briesch, Riley-Tillman, & Schlientz, 2007; Merrell, 2007). Briesch et al. (2007) suggested that these effects as well as others, such as the generosity effect, can influence ratings, and thus result in an unfair bias for or against the student being observed. They also stated that the dual role of the teacher (teaching and rating simultaneously) might adversely affect the accuracy of the data being collected on a student or may change the way the teacher interacts with the student(s) being observed, changing the outcome of the data being collected. Schlientz et al. (2009) also suggested that rater error could be due to the fact that teachers often have a difficult time focusing on any one child for an extended period of time. DBRs may be more influenced by a rater's perception of the behavior problem given that one recording is required relative to frequent recordings of actual behavior during interval recoding methods. This may explain why there is moderate correspondence between DBR and SDO recordings of academic engagement or disruptive behaviors. DBRs are still very new to the field and more research needs to be completed to examine these different possible faults, as well as to whether DBRs should be used simply as a screening device or if it should be used for high-stake decision making.

Peer Comparison Observations

While behavior observations are usually initiated due to a teacher's concern with an individual's problem behaviors, Rapport (2005) recommended that peer comparisons be used when doing direct observations to provide a type of normative data of behavior

occurrence of others within a specific natural environment. Often, normative data that does not include classroom peers provides too large of an age gap (6-12 years), or provides an expected behavior goal given a certain type of environment. Bell and Barnett (1999) stated the importance of using either local or peer micronorms when doing a direct observation to determine if behavior problems are target students struggling within the classroom or if it is a classwide problem due to the ineffective structure of the classroom management system. Peer micronorms involve using the behaviors of those children in the same classroom as the target child to compare if individual behaviors differ from group behaviors measured under the same environmental conditions (Bell & Barnett, 1999). Unfortunately, there are few studies that provide guidelines on how to choose the way in which peers are selected, such as random selection, teacher selection, purposeful selection, or for the number of peers that should be chosen (Bell & Barnet, 1999). The use of peer comparison observations has been recommended in several systematic observation coding systems, but all vary in the peer selection observation process and few report psychometric data supporting the methods (Crone & Horner, 2003).

Several studies have used peer comparisons to validate the utility of an observational system to identify students with a disorder by examining the extent that behavior observations obtained by the instruments differentiates between students with a disorder and typical developing students. For example, Carroll, Houghton, Taylor, West, and List-Kerz (2006) used peer comparisons to investigate the utility of a new instrument to measure the differences in reactivity between children with ADHD and those without ADHD. Twenty-nine pairs of teen-aged students with and without ADHD were observed

for 40-minute sessions. A child with ADHD was observed for the first minute, while one comparison student without ADHD was observed for the second minute. Observation results showed that those children with ADHD differed significantly from their peers without ADHD in solitary and interactional off-task behaviors and challenging behaviors. Inter-rater reliability conducted for 15% of the sessions was 89%. A study conducted by Forness and Esveldt (1975a) explored differences in classroom behaviors between 24 male first or second graders who were referred by the UCLA Child Psychiatry for learning and behavior problems with comparison peers. During math and reading groups, the participants' behaviors were compared to a same gender peer who exhibited no learning or behavior problems. Students were observed for six days using six second partial interval sampling. Results of this study indicated that the hospital-referred children had significantly less on-task behavior with more teacher instructional prompting, but did not exhibit significantly different problem behavior than their peers. These results indicate some utility for peer comparison observations as a screening tool in the classroom to identify individuals exhibiting behavior problems different than others in the same context.

A few researchers have investigated the effect of behavioral management on behaviors of referred students and peers due to behavior problems. For example, Forness and Esveldt (1975b) used peer comparison to look at kindergarten students to predict which of the students would continue to struggle and the degree to which in-class observations could be used as a tool to determine which students may require early intervention. Each student in four classrooms was observed for 10 days in the fall and

winter each day using a 6-second interval recording system. Each student's on-task, disruptive, and attending behavior was recorded for one interval until all students were observed and the rotation was repeated until all students were observed for one minute (i.e., ten 6-sec intervals) each day as students participated in a teacher directed group activity. Interestingly, results indicated that the identification of a child with a behavior problem might be a function of classwide behavior. That is, more students were identified as at-risk students in classrooms with the most off-task behavior when identification was based on off-task behaviors at one standard deviation or more below the average behavior of all four classrooms.

Deno (1980) conducted peer comparison observations in a study to determine whether or not excessive behaviors shown by target students were consistent with the peer norms. Ten students who were nominated by teachers as being poorly socially adjusted and 10 socially adjusted nominated peers were observed on five different occasions using a frequency recording method. Results showed that those students recommended as being poorly socially adjusted by the teacher did show deviant behaviors, but these students were frequently consistent with the peer norm of that classroom showing that the problem was not the target students, but rather the classroom management program as a whole. Based on this finding, the authors recommended that multiple observations of the target student and peers be completed to avoid making errors in recommendations due to perceived differences in behavior seen in the target student versus the classroom as a whole.

Peer comparisons are a necessary feature when observing children for evaluations

for special education classification and services to ensure that their behavior is truly deviant from their peers in a well behavior-managed classroom environment. For example, Skiba, McLeskey, Waldron, and Grizzle (1993) investigated the effectiveness of instruction and classroom management variables across classrooms as well as possible factors of the referral of children to special education on student behavior. Students' academic engaged time, inappropriate behaviors, and some behavior management strategies used by their teachers were recorded using the Code for Instructional Structure and Student Academic Response (CISSAR; Stanley & Greenwood, 1980) that employed a momentary time sampling observation with 10-second intervals. Twenty-four teachers nominated a male student from each of the five classes they believed were difficult to teach either for behavioral or academic reasons, or were showing more behavior problems than their peers. Teachers also selected three male comparison peers that they believed were not at-risk for any of these things. One of these pairs of peers was selected to be used in the study as the peer comparison. Although at-risk and not-at-risk rated students received the same assigned tasks and instructional structure, the authors found that those 24 children nominated as at-risk showed significantly more problem behavior than their selected not-at-risk peers, despite the findings that the at-risk students also received higher quality behavior management than not-at-risk students. That is, the at-risk children observed in this study received brief transitions with clear transition direction, worked in organized, fast paced classrooms, and received consistent follow through with warnings and consequences following behaviors in large group instruction.

Purpose of Current Study

Given that off-task and disruptive behaviors interfere with learning, direct observation of these external behavior problems in the classroom environment is an assessment tool frequently used to determine when and why these problems occur for some students in the classroom (Riley-Tillman, Chafouleas et al., 2009). A primary outcome of this assessment is to develop a hypothesis that will lead to effective interventions. For example, an individual behavior problem may be due to poor classroom management (classwide problem) that requires a classwide intervention, or due to individual behavioral functions, such as problem behavior, including successfully contacting peer attention, teacher attention, or escape from aversive tasks (individual problems) that requires individual intervention. Peer-comparison direct observation within the classroom environment is one way to rule out classwide problems before a focus on individual problems.

A review of the literature suggests several types of SDOs that can be used to collect peer-comparison, but few studies have validated the most accurate method to collect peer comparison data to estimate classwide behavior problems (Riley-Tillman & Burns, 2009). As such, it is critical that additional research is conducted to validate classwide assessment methods that can be used to identify and analyze types of behavior problems. An ideal method is one that is easy to learn and conduct and one that provides useful information to problem solving at the individual or classwide level. Preliminary information suggests that DBR is a promising observation method to gather useful information about classwide behavior to determine whether or not a classwide

intervention is needed. For example, DBRs have good reliability and validity when completed by different people, but have been shown to have even better reliability and validity when completed by the same person (Chafouleas et al., 2010; Christ et al., 2010). DBRs have been used successfully to observe typically developing students, as well as children with behavior problems typically seen in a classroom (Chafouleas et al., 2005; Riley-Tillman et al., 2008; Riley-Tillman, Chafouleas et al., 2009; Schlientz et al., 2009) with preliminary support for assessing classwide behavior (Riley-Tillman, Chafouleas et al., 2009). Moreover, teacher estimation of classwide behaviors using DBRs may be similar to the more complex methods such as interval or momentary time sampling procedures. However, additional research is needed to determine the relation of DBRs compared to momentary time sampling procedures, given its higher degree of accuracy in estimating individual behavior relative to other interval recording systems (Hintze, 2004; Green et al., 1982; Salvia & Hughes, 1990). Thus, the purpose of the present study was to expand the concurrent validation of DBRs conducted by Chafouleas et al. (2005) and Riley-Tillman et al. (2008), by investigating the concurrent validity of DBRs for estimating classwide behavior. The present study was designed to assess the agreement between behavior data collected using two types of peer comparison SDO procedures and DBRs. Specific research questions were:

1. What is the relationship between the estimated percentage of classwide (a) on-task behavior; and (b) disruptive behavior obtained using SDO with momentary time sampling using three peers, SDO with momentary time sampling using all peers and a Direct Behavior Rating Scale?

2. Is there a significant difference between the estimated percentage of classwide (a) on-task behavior; and (b) disruptive behavior obtained using three different classroom observation methods: momentary time sampling using three peers, momentary time sampling using Classwide of all peers, and a Direct Behavior Rating scale?

CHAPTER III

METHODS

Videotaped Students and Setting

College students were recruited from one western public university participate in video-taped classroom lesson sessions. A demographics form (see Appendix A) collected information from participating college students ($N = 113$) in nine different undergraduate psychology classes at the University. Results are presented in Table 1.

Sessions were taped and later used for observations in six classrooms at the University during different psychology lab classes. Student desks in each classroom were arranged in rows with a total of one to eight desks in a row and one to ten desks in a column depending upon the classroom. The course or lab teacher taught class as usual, while the author recorded the class via video from either the front or the back of the classroom. The course or lab leader had access to typical teaching materials (e.g., board, overhead) at the front of the room. The students' desks faced the board and the teacher. Class sizes ranged from 14 to 27 students.

Table 1

Demographics of Participants

Demographic		*n*	Percentage
Gender	Male	33	29.2%
	Female	80	70.8%
Ethnicity	Caucasian	107	94.7%
	Latino/a	3	2.7%
	Asian American	2	1.8%
	Native American	1	.9%

Materials

Classroom instruction, with the students from each class, was videotaped to allow ratings of a sample of behaviors using various observational recording strategies. A camera was placed to view all participants from the back or front of the classroom given that these are areas in the classroom where a classroom observer would usually place himself or herself. The camera was used to tape the participants as they sit through their typical 1-hour lab or lecture in the area of Psychology. Tapes were edited into 7-minute sessions if a class was taped for more than 7 minutes. This is based on studies indicating that 15 minutes is a common duration for an observation session (Chafouleas et al., 2007; Hintze & Matthews, 2004; Shapiro & Heick, 2004). Given that peers would be observed for part of that 15-minute session for a typical classroom observation, a 7-minute session was selected as a typical amount of time that peers would be observed for a peer comparison. There were a total of nine classes taped with a total of twenty-five 7-minute edited sessions that viewed a range of 14 to 27 students who were coded. An auditory marker was also inserted every 10 seconds on each of the 25 video sessions.

Dependent Measures

Two dependent variables, on-task and disruptive behaviors, were observed. These two behaviors were selected given that change in these behaviors is the most frequent intervention goal for students with emotional and behavioral disorders in the classroom (Wolraich, Feurer, Hannah, Baumgaertel, & Pinnock, 1998) and that these were most used in prior studies examining the psychometrics of DBRs (Chafouleas et al., 2010;

Chafouleas, Kilgus et al. 2009; Chafouleas et al., 2005; Riley-Tillman et al., 2008; Riley-Tillman et al., 2009). Similar to the definition used in research studies using DBRs (Chafouleas et al., 2005), on-task behavior was defined as having the student be oriented toward the teacher and engaged during the lesson or activity. Examples of on-task behaviors included sitting in their assigned seat, facing forward, participating in question/answer activity, working quietly on worksheet activities when asked, actively attending to an academic task or materials (e.g., reading, writing, using materials, working in a group), making appropriate motor responses (writing, reading out loud), and seeking appropriate teacher assistance (hand raising, asking questions; Shapiro & Heick, 2004; Walker & Severson, 1992). Also, similar to the definition used in research studies using DBRs (Chafouleas et al., 2005), student disruptive behavior was defined as behavior that is disrupting to the teacher or others in the classroom. Examples included touching, aggression, playing, making noise, out of seat, talking-out to peers and teacher without teacher permission, using profanity or sexually-related language, leaving their desk during instruction, making distracting facial expressions or obscene hand gestures to others in the classroom, and making repeated audible noises with tangible items (e.g., tapping pencil repetitiously on desk). Behaviors that were off-task but not disruptive were not specifically coded. The total percent of disruptive behavior and on-task behavior was calculated from the different measures as described in the following sections.

Independent Variables

Three different observation systems were the independent variables in this study. A

description of the Direct Behavior Rating, Three-Peer Systematic Direct Observation, and Classwide Systematic Direct Observation follows.

Direct Behavior Rating (DBR)

Two direct-behavior rating scales were used to measure on-task and disruptive behavior of each classroom (see Appendix B). The definition previously mentioned was used for each rating. Each scale was a horizontal line with vertical markings at 10 equal gradients. The gradients were marked with three quantitative anchors, 0%, 50%, and 100% at the first, middle, and end gradient mark, respectively. Raters were asked to estimate the percentage of students that exhibited on-task or disruptive behavior by marking an "X" along the continuous line on the two scales. For example, immediately after observing a lesson, the rater would place an "X" at the 80% gradient marker for on-task behavior scale when estimating that 80% of the students were on-task for most of the time.

Systematic Direct Observation (SDO)

Two peer comparison momentary sampling procedures were used in this study. One procedure used three-peer comparison students, randomly chosen, and the other procedure sampled the entire classroom during the session. Below are the sections to describe the two specific procedures that were used to record behavior on a recording form (see Appendix C and D) and the calculation of on-task and disruptive behaviors for each peer observation method.

Three-Peer Systematic Direct Observation. On-task and disruptive behavior

was measured using a 10-second momentary interval time sampling procedure to obtain estimates of behavior rate and duration. Momentary time sampling procedures have been shown to be more accurate than either whole or partial interval sampling (Hintze, 2004). Ten-second intervals were selected because of the increased accuracy of using short interval momentary time-sample (Salvia & Hughes, 1990). To improve reliability of the time sampling procedure, an audio beep was dubbed to the prerecorded tape at the end of the 10-second interval to prompt the observer recording in the correct time sampling of 10-second intervals. At the end of a 10-second interval, a trained observer looked at the student to be observed, silently count to one second, record the student as "on-task" if the student had been looking at or completing the assigned task, working quietly, and seated during the observation interval, or record the behavior as "disruptive" if the student's observed behavior met the above disruptive behavior definition. After recording each interval, observers maintained their gaze on the recording sheet until an audio beep signaled the next observation instant of a student on-task or disruptive behavior.

For the three-peers momentary time sampling method, three comparison peers were chosen by randomly selecting three seat numbers to observe for each observation session tape. These seat numbers were randomly selected from a random number generating program in Microsoft Excel. The observers of a tape were given the seat numbers of the three peers that were to be observed. The observer then began to observe and record behavior of the assigned peers in numerical order on a Three-Peer Momentary Time Sampling Recording Form (see Appendix C) for the first, second, and third 10-second interval, respectively. The observers kept repeating this rotation pattern such that

each student would always be observed in that order until the end of the session. These data were reported as the combined behavior of all three students by calculating the percentage of intervals of recorded disruptive classroom behavior and on-task behavior recorded during all intervals on the recording sheet per observation session.

Classwide Systematic Direct Observation. In this peer-observation technique, each student in the room was observed using an organized procedure and each student's behavior was recorded on a Classwide Momentary Time Sampling Recording Form (see Appendix D). Each desk in the classroom row was assigned a specific number in numerical sequence starting with desk number 1 until all desks were numbered. An observer then observed and recorded behavior of the student in desk number one in the first interval, observed and recorded the behavior of student in desk number two in the second interval, and so on. When all the behaviors of the children in the class were recorded, the observer started again on this number sequenced rotation.

Procedure

Recruitment of Students for Taping

Students were recruited from university psychology classes that offered the opportunity to receive course credit in either this research project or in various other types of opportunities during the beginning of a semester. The opportunity was announced in class as in accordance with all other lab or extra credit opportunities. At the beginning of the semester, lab classes were visited by a graduate research assistant (RA). During this visit, the RA provided a brief oral explanation of the purpose of the study,

distributed the informed consent form (two copies), one for the student records and one for the PI (see Appendix E), and provided an opportunity to answer questions. After consent was obtained, students were videotaped for a 1-hour lab session (or for the length that the lab ran). If a student in a classroom chose to not participate, they were told the camera view would not include them. However, there were no students that refused to be taped for the study.

Classroom Simulated Videotaping

A session was taped if at least 14 students from the class agreed to a videotape session during their class. During the class hour, all students were asked to sit in one of the desks that could be seen in the view of the camera. The lab instructor was present at each taping and taught the lesson as usual. The lessons consisted of teacher-directed lecture on a topic, group discussion, activity, or some combination of the above. Twenty-five taped sessions were included in the study in classes that had at least 14 students within the view of tape for seven minutes. Nineteen of the 25 videos included both lecture and class discussion. Six of the videos were class-activity only. To ensure variable ranges of classroom behaviors, different types of classrooms were taped on multiple days.

Observation Training, Data Collection, and Inter-Observer Agreement

Three psychology graduate students and the primary researcher assisted in collecting data for the study by observing student on-task and disruptive behavior while watching a taped session using a Direct Behavior Rating Form (see Appendix B) or the Systematic Momentary Time Sampling Behavior Recording Forms (see Appendix C and

D). The primary researcher trained the data collection assistants in the three observation procedures. Prior to data collection, a three step-training module (Chafouleas, Riley-Tillman, & Christ; directbehaviorratings.com) was used to train the RAs on the Direct Behavior Rating scale. The first step taught the DBR form and the behavior definitions. The second step taught how to rate academic engagement (on-task behavior in this current study) and disruptive behavior using multiple video clips of a classroom. The third step required the trainee to practice DBR ratings with feedback on rating accuracy while viewing multiple videos. A completion certificate that is mailed to trainees was required before a research assistant is considered trained to code on DBR for this study.

For the momentary time sampling procedures, the observers received both verbal and written instructions on how to properly record behavior on the Momentary Time Sampling Forms, which was followed by modeling of the procedure from the trainer. Practice sessions were conducted on tapes that had been previously viewed and recorded using momentary sampling of students' on-task and disruptive behaviors (as described in the systematic observation section). At the end of practice sessions, inter-observer agreement (IOA) for each observed on-task and disruptive event were calculated on an interval by interval point basis, namely agreement steps (in which both observers agree that a behavior did or did not occur) divided by agreements plus disagreements with the remainder multiplied by 100%. Each trainee reached IOA agreement of over 80% for three consecutive sessions to be considered trained.

Once training of observers was completed, data for this study were collected by having the observers record the on-task and disruptive behavior, while viewing the

simulated classroom videotapes. For each session, the direct behavior rating scale was coded first to decrease influence results of a momentary time sampling procedure. Then, the classwide momentary systematic direct observation or the three-peer momentary systematic direct observation was counterbalanced directly following the DBR.

During the study, inter-observer agreement was evaluated by two independent trained observers for 100% of the taped observation sessions for all three types of observation methods used in this study. Inter-rater reliability for the DBR method was evaluated using Pearson product-moment correlation coefficients. There was a strong correlation between two observer ratings the DBR-on (r = .93) and DBR-disruptive (r = .98). Inter-rater reliability for the two momentary sampling methods was evaluated by calculating Kappa coefficients between scorer 1 and scorer 2 coding on either the disruptive or on-task behavior coded interval by interval. The results of the inter-rater analysis are Kappa = 0.93 for the three-peer method and Kappa = 0.88 for the classwide method. Kappas from 0.40 to 0.59 are considered as moderate, 0.60 to 0.79 as substantial, and 0.80 as outstanding (Landis & Koch, 1977) level of agreement between two sets of dichotomous ratings or scores.

CHAPTER IV

RESULTS

Several statistical and descriptive procedures were used to investigate the experimental questions. First, descriptive means, standard deviations and ranges obtained for each of the two peer monitoring systems as well as the direct behavior rating percentages are presented in Table 2.

Second, Pearson product-moment correlation coefficients were calculated to examine agreement between systematic direct observation scores using three peers, systematic direct observation data using all peers, and DBRs of both on-task and disruptive behaviors. Data for all three methods are also graphed in Figures 1 and 2. Correlation coefficients between DBR and the three-peer and classwide random momentary time sampling estimates for on-task behavior and disruptive behavior are presented in Tables 3 and 4. Ratings of the on-task behaviors using the DBR and SDO were significantly correlated. Specifically, there was a strong relationship between DBR and the SDO classwide on-task estimates. A total of 37% of the variance in the SDO classwide data was consistent with the DBR data (Cohen, 1988; Crocker & Algina, 1986). There was not a significant correlation between the DBR and the three-peer observation data, nor between the three-peer data and the classwide observation data. There was a moderate relationship between the on-task DBR and three-peer on-task with 13% of the variance in the SDO data as a portion of the DBR data.

For the disruptive behavior estimate, there was a significant strong correlation between DBR and both of the SDO methods with 43% for the classwide SDO and 39%

of the three-peers SDO variance consistent with the DBR data.

Table 2

Descriptive Statistics for On-Task and Disruptive Recorded Estimates

		Observation Method		
Behavior		DBR	Three-peer	Classwide
On-Task	Mean	73.28%	71.56%	70.12%
	SD	18.30	11.61	10.96
	Range	25 to 95	48 to 98	52 to 90
	Kurtosis	.59	.01	-.90
	Std. Error of Kertosis	.902	.90	.90
	Skewness	-1.09	.27	-.38
	Std. Error of Skewness	.46	.46	.46
Disruptive	Mean	18.88%	18.21%	18.23%
	SD	15.94	11.01	9.92
	Range	2 to 65	0 to 45	5 to 40
	Kurtosis	1.84	-.154	-.076
	Std. Error of Kertosis	.90	.90	.90
	Skewness	4.51	.59	.70
	Std. Error of Skewness	.46	.46	.46

Between the two SDO measures, there was a significant strong correlation with 30% of the variance explained. Overall, the strongest correlations were between the DBR and classwide data for both on-task and disruptive behaviors.

Table 3

Correlations Between Observation Methods for the On-Task Behavior Estimate

	Classwide SDO On-task	Three-Peer SDO On-Task
DBR On	.609**	.367
Three-Peer SDO On-task	.498	

Note. DBR = Direct Behavior Ratings and SDO = Systematic Direct Observation
*p<.05, **p<.01

Table 4

Correlations Between Observation Methods for the Disruptive Behavior Estimate

	Classwide SDO Disruptive	Three-Peer SDO Disruptive
DBR Disruptive	.658**	.624**
Three-Peer SDO Disruptive	.546**	

Note DBR = Direct Behavior Ratings and SDO = Systematic Direct Observation
*p<.05, **p<.01

Third, independent *t* tests were conducted to compare the effect of coding behaviors using the SDO first or second after the DBR evaluation for the classwide and three-peer methods. There was no significant difference between the two orders for on-task for classwide, t (23)= .23, p = .82, and three-peers, t (23)= .1.42, p = .17.

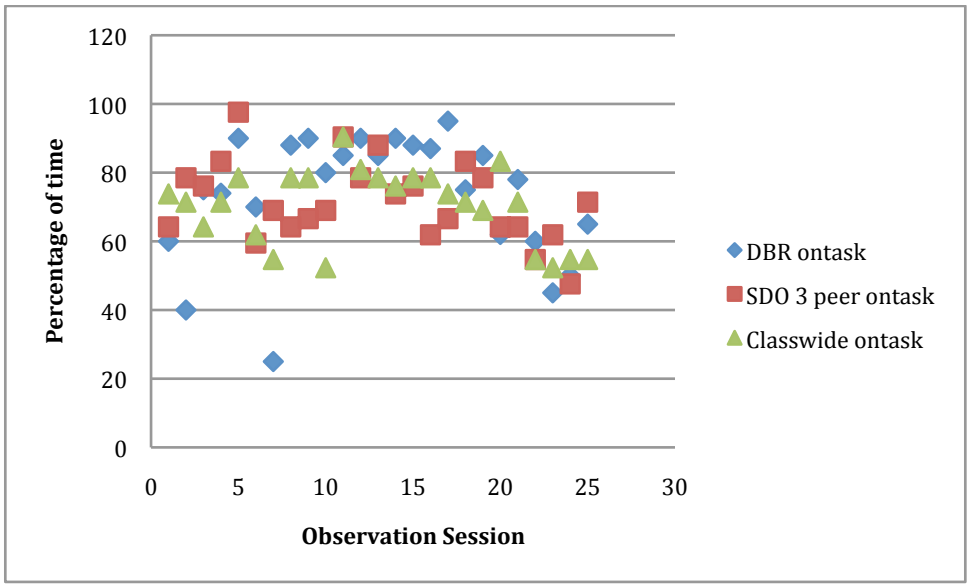


Figure 1. Percentages of coded on-task behavior.

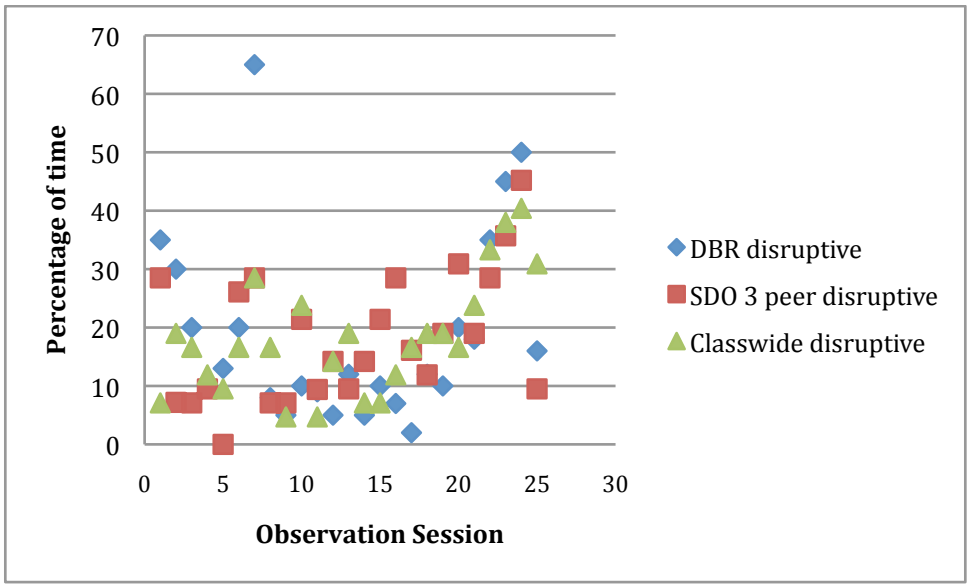


Figure 2. Percentages of coded disruptive behavior.

Likewise, there was not a significant difference between the disruptive estimates for classwide, t (23)= .46, p = .65, and three-peer, t (23)= .28, p = .78, which indicates no ordering effect for both SDO methods for both behaviors.

Finally, two separate univariate repeated-measure analyses of variance (ANOVA) were conducted to evaluate differences between the three observation methods (DBR, SDO with three peers, SDO with entire class) on the on-task and the disruptive score dependent measures. The repeated-measure ANOVA results showed that there was not a significant difference of on-task behavior among the three methods used, Wilks' Lamba = .942, F (2, 23) = .71, p = .50. There was also no significant effect for the type of method used on disruptive behavior, Wilks' Lamba = .942, F (2, 23) = 70, p = .50. Effect sizes from the ANOVAs, indicated by partial η^2 (i.e., percent of variance for which the effect accounted) was .06 for both on-task and disruptive behavior. Thus, on-task and disruptive behavior showed a moderate estimate of effect size, respectively, based on Cohen's (1988) criteria for evaluating effect sizes in ANOVA: small 1% to 5.8% of variance; medium 5.9% to 13.7% of variance; and large 13.8% of variance.

Although there were no significant differences between the three measures, it is important to note that the percentages were not identical between the three methods. Table 5 presents the mean differences and ranges between the various pairs of observation assessment methods for on-task and disruptive behaviors. Although mean differences were low, results show that differences between SDO and DBR phase means ranged between a 15% difference to a larger 44% difference.

Given that a purpose of this study was to further explore the utility of DBR in a problem-solving model to determine if there is a classwide behavior management problem that would require classwide intervention before further assessing an individual

problem, the frequency of each percentage obtained for the on-task and disruptive behaviors per session are presented in Table 6.

Table 5

Mean Differences and Ranges Between the Observation Methods

Differences between methods		Minimum	Maximum	Mean	Standard. Deviation
On-task	DBR and 3 peer	-44.00	28.40	1.7160	17.71399
	DB and CW	-31.40	27.70	3.1440	14.51660
	3 peer and CW	-19.10	19.10	1.4280	11.32713
Disruptive	DB and CW	-14.90	36.50	.6440	12.02106
	DBR and 3 peer	-21.50	36.50	.6604	12.50584
	3 peer and CW	-21.40	21.40	-.0164	10.01255

Benchmarks were also used to determine at-risk and no-risk status for classwide behavior management problems. The benchmarks were based on multiple studies examining average classroom on-task behaviors that have reported on-task ranges from 77% to 86% (Forness & Esveldt, 1975b; Hintze & Matthews, 2004; Shin, Ramsey, Walker, Stieber, & O'Neill, 1987; Skiba et al., 1993). Specifically, a class was judged as a no-risk class if the on-task behaviors were at or above 80% and the disruptive behaviors was at or lower than 20%. Cohen's Kappa statistics were calculated measuring the agreement of at-risk and no-risk judgments between DBR and the two SDO methods. Results of this analysis for on-task behaviors were $\kappa = .098$ for the classwide SDO and κ

= .092 for the three-peer SDO. Results of this analysis for disruptive behaviors were κ = .481 for the classwide SDO, κ = .464 for the three-peer SDO. The on-task behavior can be characterized as a slight agreement and the disruption behavior as a moderate agreement (Landis & Koch, 1977).

Table 6

Frequency for Coded Percentages for At-risk and No-Risk Classwide Behavior Management Problems Outcomes

Categorical Outcome		Coded percent	Frequency		
			DBR	SDO 3 peer	SDO CW
On task	At-risk	0% to 9%	0	0	0
		10% to 19%	0	0	0
		20% to 29 %	1	0	0
		30% to 39%	0	0	0
		40% to 49%	2	1	0
		50% to 59%	1	2	6
		60% to 69%	4	10	3
		70% to 79%	5	7	13
	No-risk	80% to 89%	7	3	2
		90% to 100%	5	2	1
Disruptive	At-risk	90% to 100%	0	0	0
		80% to 89%	0	0	0
		70% to 79%	0	0	0
		60% to 69%	1	0	0
		50% to 59%	1	0	0
		40% to 49%	1	1	1
		30% to 39%	3	2	3
		21% to 29%	0	7	3
	No-risk	10% to 20%	12	6	12
		0% to 9%	7	9	6

CHAPTER V

DISCUSSION

Results of the current study add to the existing, but limited amount of evidence indicating that a Direct Behavior Rating form measuring on-task and disruptive behavior has potential to be used as a screening tool to collect peer comparison data to estimate classwide behavior problems. Similarly to DBR observations for individuals (Riley-Tillman, Methe et al., 2009), a moderate degree of agreement was found between DBR ratings and SDO methods. In this study, the DBR estimates were more strongly related to the classwide SDO estimates as compared to the three-peer SDO estimates. Interestingly, the relationship between the DBR and classwide SDO was stronger than the relationship between the two SDO's on both observed behaviors.

This investigation also had similar findings to those shown from Chafouleas et al. (2005) and Riley-Tillman et al. (2008), which showed a moderate to strong degree of similarity between raters, ranging between $r = .48$ to $r = .87$ on the SDOs and DBRs for on-task and disruptive behaviors. However, it is important to note that the two prior studies used a DBR format with a likert-scale with a 0-5 range and used 20-second intervals to collect SDO data while this study used a 0-10 likert range on the DBR and 10-second intervals as the SDO methods. Given that Riley-Tillman et al. (2008) and Schlientz et al. (2009) found that the training resulted in higher reliability between raters, it is also important to note that this study employed similar intensive training methods (verbal and written explanations, modeling, practice on tapes, and completion of the DBR training program that resulted in greater reliability). Moreover, the observers needed to

match with the master coding sheet with at least 90% agreement on three separate occasions to be considered trained in the Riley-Tillman studies, while this study required only 80% agreement on three separate occasions. Thus, intensive training on DBR is likely needed to get similar results.

Christ et al. (2011) and Riley-Tillman, Chafouleas et al. (2009) showed that data from DBRs tended to result in a slightly more negative behavioral outcome of the student than did the SDO data. This outcome differed slightly from the study completed by Riley-Tillman et al. (2011), which showed a positive bias for both on-task and disruptive behavior on the DBR compared to the SDO. In comparison, results from this current study found that 6 of the 25 (24%) of the DBR on-task ratings were lower than the SDO classwide on-task ratings and 14 of the 25 (56%) of the DBR disruptive ratings were higher than the SDO classwide disruptive behavior ratings. Thus, on-task tended to result in a more positive behavioral outcome whereas the disruptive was more evenly distributed between a negative and positive outcome relative to the SDO outcomes. Understanding the degree that error, rater perceptions, or other factors may explain the observed bias.

More specifically, these results extended findings on classwide DBR assessments. Prior to this study, Riley-Tillman, Methe et al. (2009) explored the utility for DBR to identify effective classwide behavior change with intervention in a case study. The results from this Riley-Tillman study showed high agreement between DBR and SDO momentary-time sampling data points (κ range between .65 to .80) when examining agreement between identification of three levels (moderate, large, and very large) of

classwide response to intervention set at 55%, 60%, and 65% benchmarks respectively. The purpose of this study was to further explore the utility of DBR in a problem-solving model to determine if there is a classwide behavior management problem that would require classwide intervention before further assessing an individual problem. Although Salvia et al. (2009) suggested that all classroom observation methods summarize data collected using standardized rules, clear benchmarks for selected behaviors and a definition of what constitutes a well managed classroom, it is not yet well-defined in the literature (Hintze & Matthews, 2004). The benchmark in this study was based on a few studies that suggested 80% on-task behavior is typically observed in general educations classrooms across tasks (Forness & Esveldt, 1975a; Hintze & Matthews, 2004; Shin et al., 1987; Skiba et al., 1993). When determining the dichotomous at-risk and no risk existence of a classwide behavior management problem based on the benchmark criteria used in this paper, the on-task and disruptive behaviors only showed a slight to moderate agreement with SDO methods (Landis & Koch, 1977). Furthermore, a higher percentage of disruption DBR judgments corresponded to the SDO risk judgments than on-task DBR risk judgments. Specifically, 80% (n = 20) and 76% (n = 19) of DBR decisions based on disruptive behaviors were the same as the SDO classwide and three-peer method. For on-task behavior estimates, 56% (n = 14) of the DBR decisions (at-risk or no-risk) were the same as the three and classwide SDO methods. A class was rated as a no-risk on the on-task DBR ratings but identified as an at-risk class on 40% (n = 10) and 36% (n = 9) of the sessions on the classwide and three-peer SDO. Thus, identified no risk DBR classes may be under-identified as being at-risk relative to the SDO estimates for on-task behavior.

Considering that the implication of classwide problems for many students in the class, this finding provides cautious findings when using DBR as a screener in regard to decision making of behavioral problems at the classwide level before assessing the level of an individual child. These results suggest that the classwide DBR data may not be as robust across classrooms or when estimating classroom management using an 80% on-task or 20% disruption benchmark.

Limitations

While the results from the current study provide additional information on prior and current research on DBRs and SDOs, there are some limitations to be noted. First, this study used undergraduate students as the unit of observation instead of elementary-aged students due to the limitation of getting parental and school permission. Although the inclusion college students allowed a review of a large group of student behavior in a natural classroom setting, the external validity of the findings from this study is limited to classrooms in elementary and secondary settings and should be interpreted with caution. Moreover, taped sessions did not allow the observation of a typical class size that would consist of more than 14 students. However, it is not known to what degree a larger class size or how many disruptive students would vary or skew results on any observation measure. Finally, the fewer classrooms that fell at the lower percentages of on-task behaviors and higher percentages of disruption within the no-risk than at-risk category may have influenced results.

Further Directions

Clearly, these preliminary results suggest a need for further research on the most optimal DBR procedures for making accurate educational decisions at the classwide level. Future studies should systematically replicate classwide DBR estimates of behavior across important target behaviors with standardized definitions in the classroom settings that accurately differentiate between good and poor behavior managed classrooms. Furthermore, given that one to three behaviors are listed on DBRs used by prior researchers, additional investigations exploring how many behaviors can be rated using a DBR, while still remaining accurate may be warranted when making judgments about effective classwide management. Differences in the screening potential of DBRs with different number of students in a class and different observation times should also be further explored. There is currently no literature suggesting a minimum, nor maximum needed and class sizes differ widely between school, programs, and districts. Furthermore, knowledge on how many DBRs should be completed to accurately determine if there is a classwide behavior problem is needed. It would be necessary to determine if one is sufficient, or if multiple ones are needed for decision-making. Lastly, more research on the type of training/duration/intensity that teachers or other observers would need to accurately rate behaviors on a DBR is needed to expand the literature in this area.

Conclusion

In sum, these preliminary results indicate that DBRs and SDOs data correlate well

with each other, which suggests that the DBR has potential to be used as a screener or estimation tool for classwide behavior problems. Additional research is warranted to confirm these results given that DBR may be a more feasible screening option for judging effective classwide behavior management systems.

REFERENCES

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (Revised 4th ed.). Washington, DC: Author.

Bell, S.H., & Barnett, D.W. (1999). Peer micronorms in the assessment of young children: Methodological review and examples. *Topics in Early Childhood Special Education, 19*(2)*,*112-123.

Bowen, J.M., Jenson, W.R., & Clark, E. (2004). *School-based interventions for students with behavior problems*. New York, NY: Springer.

Bramlett, R.K., Murphy, J.J., Johnson, J., Wallingsford, L., & Hall, J.D. (2002). Contemporary practices in school psychology: A national survey of roles and referral problems. *Psychology in the Schools, 39*, 327-335.

Briesch, A.M., Riley-Tillman, T.C., & Schlientz, M. (2007). *Project VIABLE technical report: A review of sources of effort in social behavior assessment*. Storrs: University of Connecticut.

Campbell, S.B. (1995). Behavior problems in preschool children: A review of recent research. *Journal of Child Psychology and Psychiatry*, *36,* 113–149.

Carroll, A., Houghton, S., Taylor, M., West, J., & List-Kerz, M. (2006). Responses to interpersonal and physically provoking situations: The utility and application of an observation schedule for school-aged students with and without attention deficit/hyperactivity disorder. *Educational Psychology, 26*(4), 483-498.

Chafouleas, S.M., Briesch, A.M., Riley-Tillman, T.C., Christ, T.C., Black, A.C., & Kilgus, S.P. (2010). An investigation of the generalizability and dependability of

direct behavior rating single item scales (DBR-SIS) to measure academic engagement and disruptive behavior of middle school students. *Journal of School Psychology, 48*(3), 219-246.

Chafouleas, S.M., Christ, T.J., & Riley-Tillman, T.C. (2009). Generalizability of scale gradients on direct behavior ratings. *Educational and Psychological Measurement*, *69*(1), 157-173.

Chafouleas, S.M., Kilgus, S.P., & Hernandez, P. (2009). Using direct behavior rating (DBR) to screen for school social risk. *Assessment for Effective Intervention, 34*(4), 214-223.

Chafouleas, S.M., McDougal, J.L., Riley-Tillman, T.*C.,* Panahon, C.J., & Hilt, A.M. (2005). What do daily behavior report cards (DBRCs) measure? An initial comparison of DBRCs with direct observation for off-task behavior. *Psychology in the Schools, 42,* 669-676.

Chafouleas, S.M., Riley-Tillman, T.C., & Christ, T.J. (2009). Direct Behavior Rating (DBR): An emerging method for assessing social behavior with a tier intervention system. *Assessment for Effective Intervention, 34*(4), 195-200.

Chafouleas, S.M., Riley-Tillman, T.C., & Sassu, K.A. (2006). Acceptability and reported use of daily behavior report cards among teachers. *Journal of Positive Behavior Interventions, 8*, 174-182.

Chafouleas, S., Riley-Tillman, T.C., & Sugai, G. (2007). *School-based behavioral assessment: Informing intervention and instruction.* New York, NY: Guilford.

Christ, T.J., Riley-Tillman, T.C., Chafouleas, S.M., & Boice, C.H. (2010).

Generalizability and dependability of Direct Behavior Rating Scales (DBRS) across raters and observations. *Educational and Psychological Measurement*, *70*(4), 1-19.

Christ, T.J., Riley-Tillman, T.C., Chafouleas, S.M., & Jaffery, R. (2011). Direct behavior rating: An evaluation of alternate definitions to assess classroom behavior. *School Psychology Review, 40*(2), 181-199.

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

Conroy, M.A., Stichter, J.P., Daunic, A., & Haydon, T. (2008). Classroom-based research in the field of emotional and behavioral disorders: Methodological issues and future research directions. *The Journal of Special Education, 41*(4), 209-222.

Cooper, J.O., Heron, T.E., & Heward, W.L. (2007). *Applied behavior analysis*. Englewood Cliffs, NJ: Prentice Hall.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Orlando, FL: Harcourt Brace.

Crone, D.A., & Horner, R.H. (2003). *Building positive behavior support systems in schools: Functional behavior assessment*. New York, NY: The Guilford.

Deno, S.L. (1980). Direct observation approach to measuring classroom behavior. *Exceptional Children, 46*(5), 396-399.

DuPaul, G.J., & Hoff, K.E. (1998). Reducing disruptive behavior in general education classrooms: The use of self-management strategies. *School Psychology Review, 27*(2), 290-303.

DuPaul, G.J., & Stoner, G. (2003). *ADHD in the schools: Assessment and intervention*

strategies (2nd ed.). New York, NY: Guilford.

Eidle, K., Truscott, S.D., Meyers, J., & Boyd, T. (1998). The role of prereferral intervention teams in the early intervention and prevention of mental health problems. *School Psychology Review, 27*, 204-216.

Ellis, J., & Magee, S. (2004). Modifications to basic functional analysis procedures in school settings: A selective review. *Behavioral Interventions, 19*(4), 205-228

Epstein, M., Atkins, M., Cullinan, D., Kutash, K., & Weaver, R. (2008). Reducing behavior problems in the elementary school classroom: A practice guide (NCEE #2008-012). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http:// ies.ed.gov/ncee/wwc/publications/practiceguides

Evertson, C.M., & Weinstein, C. S. (Eds.) (2006). *Handbook of classroom management: Research, practice, and contemporary issues*. Mahwah, NJ: Erlbaum.

Forness, S.R., & Esveldt, K.C. (1975a). Classroom observation of children with learning and behavior disorders. *Journal of Learning Disabilities*, *8*(6), 382-385.

Forness, S.R., & Esveldt, K.C. (1975b). Prediction of high-risk kindergarten children through classroom observation. *The Journal of Special Education*, *9*(4), 375-387.

Gimpel, G.A., & Holland, M.L. (2003). *Emotional and behavioral problems of young children: Effective interventions in the preschool and kindergarten years.* New York, NY: The Guilford Press.

Green, S.B., McCoy, J.F., Burns, K.P., & Smith, A.C. (1982). Accuracy of observational data with whole interval, partial interval, and momentary time sampling recording techniques. *Journal of Behavioral Assessment, 4*(2), 103-118.

Gresham, F.M., Watson, T.S., & Skinner, C.H. (2001). Functional behavior assessment: Principles, procedures, and future directions. *School Psychology Review, 30,* 156–172.

Handler, M.W., Rey, J., Connell, J., Their, K., Feinberg, A., & Putnam, R. (2007). Practical considerations in creating school-wide positive behavior supports in public schools. *Psychology in the Schools*, *44*(1), 29-39.

Hieneman, M., Dunlap, G., & Kincaid, D. (2005). Positive support strategies for students with behavior disorders in general education settings. *Psychology in the Schools 42*(8), 779-794.

Hintze, J.M. (2004). Behavioral observation: Self-constructed. In T.S. Watson & C.H. Skinner (Eds.), *Encyclopedia of School Psychology* (pp. 43-45). New York, NY: Kluwer Academic/Plenum.

Hintze, J.M., & Matthews, W.J. (2004). The generalizability of systematic direct observations across time and setting: A preliminary investigation of the psychometrics of behavior observation. *School Psychology Review, 33*(2), 258-270.

Huston-Stein, A., Friedrich-Cofer, L., & Susman, E.J. (1977). The relation of classroom structure to social behavior, imaginative plan, and self-regulation of economically disadvantaged children. *Child Development, 48,* 908-916.

Kauffman, J.M. (2001). *Characteristics of emotional and behavioral disorders of children and youth* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

Landau, S., & Swerdlik, M.E. (2005). What you see is what you get: A commentary on school-based direct observation systems. *School Psychology Review*, *34* (4), 529-536.

Landis, J.R., & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33,* 159-174.

Lavigne, J.V., Arend, R., Rosenbaum, D., Binns, H. J., Christoffel, K.K., & Gibbons, R.D. (1998). Psychiatric disorders with onset in the preschool years: Stability of diagnoses. *Journal of the American Academy of Child & Adolescent Psychiatry, 37*, 1246–1254.

Merrell, K.W. (2007). *Behavioral, social, and emotional assessment of children and adolescents* (2nd ed.). Mahwah, NJ: Routledge.

Miltenberger, R.G. (2005). Strategies for measuring behavior change. In L.M. Bambara & L. Kern (Eds.) *Individualized support for students with problem behavior: Designing positive behavior plans.* New York, NY: Guilford.

Qi, C.H., & Kaiser, A.P. (2003). Behavior problems of preschool children from low income families: Review of the literature. *Topics in early childhood special education, 23*, 188–216.

Rapport, M.D. (2005). Commentary: School-based observations of children at school: Promise with prudence. *School Psychology Review, 34*(4), 537-542.

Riley-Tillman, T.C., & Burns, M.K. (2009). *Evaluating educational interventions:*

Single-case design for measuring response to intervention. New York, NY: Guilford.

Riley-Tillman, T.C., Chafouleas, S.M., Christ, T.J., Briesch, A.M., & LeBel, T.J. (2009). The impact of item wording and behavioral specificity on the accuracy of Direct Behavior Rating (DBR). *School Psychology Quarterly*, *24*, 1-12

Riley-Tillman, T.C., Chafouleas, S.M., Sassu, K.A., Chanese, J.A.M., & Glazer, A.D. (2008). Examining the agreement of direct behavior ratings and systematic direct observation data for on-task and disruptive behavior. *Journal of Positive Behavior Interventions, 10*(2), 136-143.

Riley-Tillman, T. C., Christ, T. J., Chafouleas, S. M., Boice, C. H., & Briesch, A. M. (2011). The impact of observation duration on the accuracy of data obtained from Direct Behavior Rating (DBR). *Journal of Positive Behavior Interventions, 12*(2), 119-128.

Riley-Tillman, T.C., Methe, S.A., & Weegar, K. (2009). Examining the use of Direct Behavior Rating methodology on classwide formative assessment: A case study. *Assessment for Effective Intervention 34*, 224–230.

Salvia, J., & Hughes, C.A. (1990). *Curriculum-based assessment: Testing what is taught.* New York, NY: Macmillan.

Salvia, J., Ysseldyke, J., & Bolt, S. (2009). *Assessment in special and inclusive education* (11th ed). Belmont, CA: Cengage Learning.

Satcher, D. (1999). *Mental health: A report of the Surgeon General*. Retrieved November 10, 2008, from http://www.surgeongeneral.gov/library/mentalhealth/home.html.

Schlientz, M.D., Riley-Tillman, T.C., Briesch, A.M., Walcott, C.M., & Chafouleas, S.M. (2009). The impact of training on the accuracy of direct behavior ratings (DBRs). *School Psychology Quarterly, 24,* 73-83.

Shapiro, E.S., & Clemens, N.H. (2007). Conducting systematic direct classroom observations to define school-related problems. In R. Brown-Chidsey & J.A. Cummings (Eds.), *Assessment for intervention: A problem solving approach* (pp. 175-192). New York, NY: Guilford.

Shapiro, E.S., & Heick, P.F. (2004). School psychologist assessment practices in the evaluation of students referred for social/behavioral/emotional problems. *Psychology in the Schools, 41*(5), 551-561.

Shin, M.R., Ramsey, E., Walker, H.M., Stieber, S., & O'Neill, R.E. (1987). Antisocial behavior in school settings: Initial differences in an at-risk and normal population. *The Journal of Special Education, 21*(2), 69-84.

Simonsen, B., Fairbanks, S., Briesch, A., Myers, D., & Sugai, G. (2008). Evidence-based practices in the classroom management: Considerations for research to practice. *Education and Treatment of Children, 31*(3), 351-380.

Skiba, R.J., McLeskey, J., Waldron, N.L., & Grizzle, K. (1993). The context of failure in the primary grades: Risk factors in low and high referral rate classrooms. *School Psychology Quarterly*, *8*(2), 81-98.

Stanley, S.O., & Greenwood, C.R. (1980). *CISSAR. Code for instructional structure and student academic response.* Juniper Garden Children's Project, University of Kansas. Kansas City, KS.

Susman, E.J., Huston-Stein, A., & Friedrich-Cofer, L. (1980). Relation of conceptual tempo to social behaviors of head start children. *The Journal of Cenetic Psychology, 137,* 17-20.

Volpe, R.J., DiPerna, J.C., Hintze, J.M., & Shapiro, E.S. (2005). Observing students in classroom settings: A review of seven coding schemes. *School Psychology Review, 34*(4), 454-474.

Walker, H., & Severson, H. (1992). *Systematic Screening of Behavior Disorders (SSBD): A Multiple Gating Procedure.* Longmont, CO: Sopris West.

Wilson, M.S., & Reschly, D.J. (1996). Assessment in school psychology training and practice. *School Psychology Review*, *25*(1), 9-23.

Wolraich, M.L., Feurer, I.D., Hannah, J.N., Baumgaertel, A., & Pinnock, T.Y. (1998). Obtaining systematic teacher reports of disruptive behavior disorders utilizing DSM-IV. *Journal of Abnormal Child Psychology, 26*(2), 141-153.

APPENDICES

Appendix A

Demographic Form

Demographics Form

Gender:

- Female
- Male
- Other

Age: _____

Major: _____

College Year:

- Freshman
- Sophomore
- Junior
- Senior
- Other (please specify):_____

Ethnicity:

- Caucasian
- Middle Eastern
- Native American
- Black
- Hispanic
- Asian
- Other (please specify):_____

Appendix B

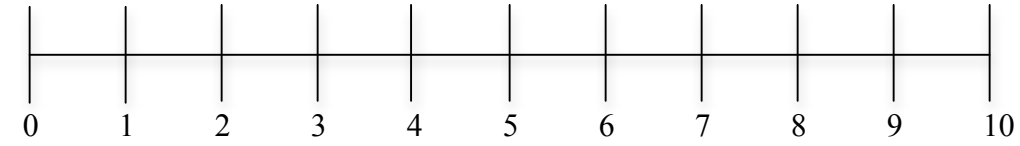
Direct Behavior Rating Form

Direct Behavior Rating Form

Directions: Place a dot on the line that best reflects the **percentage of the time the classroom exhibited the specified behavior during the observation sessions**. Specific behaviors to be rated are defined as follows:

On-task behavior: Defined as having the student oriented toward the teacher and/or is actively engaged in instructional activities (Chafouleas et al., 2005). Examples include: Sitting in assigned seat, facing forward, participating in question/answer activities, hand raising, appropriate motor responses and working quietly on assignments Shapiro, 2004; Walker & Severson, 1992.

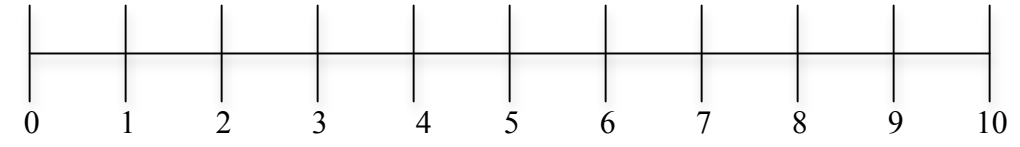
The target student was on-task ___% of the time.



0% 50% 100%

Disruptive Behavior: Behavior that is disrupting to the teacher or the classroom. The student is engaged in disruptive behavior, including touching, vocalizing, aggression, playing, disorientating, making noise, being out of seat (Kehle et al., 1986) making repeated audible noises with tangible items (pencils) making facial distractions and talking to other students.

The target student was disruptive ___% of the time.



0% 50% 100%

Appendix C

Coding Sheet for Three-Peer Observation

Coding Sheet for Three-Peer Observation

1. Complete a 7-minute observation sheet in a class with three-peer comparisons.

2. To complete the chart:

   a) Observe the appropriate child at the end of each 10-second interval (marked by audible beep).

   b) Mark the appropriate codes in BOX 1 at the end of the 10-second interval:

   ON = Defined as having the student oriented toward the teacher and/or is <u>actively</u> engaged in instructional activities (Chafouleas et al., 2005). Examples provided by Shapiro (2004) and Walker and Severson (1992) include:
   - Sitting in seat facing forward
   - Participating in question/answer activities
   - Hand raising when answering a question
   - Appropriate motor responses
   - Working quietly on assignment
   - Writing
   - Typing notes
   - Looking at person talking

   DIS = Behavior that is disrupting to the teacher or the classroom (Kehle et al., 1986). The student is engaged in disruptive behavior, including touching: vocalizing, aggression, playing, disorientating, making noise, being out of seat making repeated audible noises with tangible items (pencils) making facial distractions and talking to other students.
   - Touching another student
   - Vocalizing without raising hand
   - Aggression
   - Playing with another student or object in a distracting way
   - Being out of seat
   - Making noise with a tangible object (pencil tapping)
   - Attempts to get another student's attention (turning, calling name, tapping on shoulder)
   - Making facial distractions
   - Using cell phone, or mp3 player

   Off= Being off task, but not disruptive. This is when the student is being passively off-task
   - Reading book or newspaper
   - Not looking at teacher or taking notes
   - Looking off "into space"

c) Observe the next child at the end of the next 10-second interval and record one of the 2 behaviors in BOX 2 if it is observed at the end of the 10-second interval.

d) Complete each BOX in this manner using 10-second intervals.

1 ON DIS OFF	2 ON DIS OFF	3 ON DIS OFF	4 ON DIS OFF	5 ON DIS OFF	6 ON DIS OFF

Three-Peer Observation Form

Video Tape number: _____

Peer 1	Peer 2	Peer 3	Peer 1	Peer 2	Peer 3
1 ON DIS OFF	2 ON DIS OFF	3 ON DIS OFF	4 ON DIS OFF	5 ON DIS OFF	6 ON DIS OFF
7 ON DIS OFF	8 ON DIS OFF	9 ON DIS OFF	10 ON DIS OFF	11 ON DIS OFF	12 ON DIS OFF
13 ON DIS OFF	14 ON DIS OFF	15 ON DIS OFF	16 ON DIS OFF	17 ON DIS OFF	18 ON DIS OFF
19 ON DIS OFF	20 ON DIS OFF	21 ON DIS OFF	22 ON DIS OFF	23 ON DIS OFF	24 ON DIS OFF
25 ON DIS OFF	26 ON DIS OFF	27 ON DIS OFF	28 ON DIS OFF	29 ON DIS OFF	30 ON DIS OFF
31 ON DIS OFF	32 ON DIS OFF	33 ON DIS OFF	34 ON DIS OFF	35 ON DIS OFF	36 ON DIS OFF
37 ON DIS OFF	38 ON DIS OFF	39 ON DIS OFF	40 ON DIS OFF	41 ON DIS OFF	42 ON DIS OFF

Scoring Direct Social Interaction Observation Form

Total:

On-task: _____total marked/ _42___total boxes *100% = ____
Disruptive: _____total marked/ __42__total boxes *100% = ____

Appendix D

Coding Sheet for Classwide Observation

Coding Sheet for ClassWide Observation

1. Complete a 7-minute observation sheet in a class with Classwide comparisons.

2. To complete the chart:

   a) Observe the appropriate child at the end of each 10-second interval (marked by audible beep).

   b) Mark the appropriate codes in BOX 1 at the end of the 10-second interval:

   ON = Defined as having the student oriented toward the teacher and/or is actively engaged in instructional activities (Chafouleas et al., 2005). Examples provided by Shapiro (2004) and Walker and Severson (1992) include:
   - Sitting in seat facing forward
   - Participating in question/answer activities
   - Hand raising when answering a question
   - Appropriate motor responses
   - Working quietly on assignment
   - Writing
   - Typing notes
   - Looking at person talking

   DIS = Behavior that is disrupting to the teacher or the classroom (Kehle et al., 1986). The student is engaged in disruptive behavior, including touching: vocalizing, aggression, playing, disorientating, making noise, being out of seat making repeated audible noises with tangible items (pencils) making facial distractions and talking to other students.
   - Touching another student
   - Vocalizing without raising hand
   - Aggression
   - Playing with another student or object in a distracting way
   - Being out of seat
   - Making noise with a tangible object (pencil tapping)
   - Attempts to get another student's attention (turning, calling name, tapping on shoulder)
   - Making facial distractions
   - Using cell phone, or mp3 player

   Off= Being off task, but not disruptive. This is when the student is being passively off-task
   - Reading book or newspaper
   - Not looking at teacher or taking notes
   - Looking off "into space"

c) Observe the next child at the end of the next 10-second interval and record one of the 2 behaviors in BOX 2 if it is observed at the end of the 10-second interval.

d) Complete each BOX in this manner using 10-second intervals.

1	2	3	4	5	6
ON DIS OFF	ON DIS OFF	ON DIS OFF	ON DIS OFF	ON DIS OFF	ON DIS OFF

Classwide Peer Observation Form

Video Tape number: _____

Entire Class

1 ON DIS OFF	2 ON DIS OFF	3 ON DIS OFF	4 ON DIS OFF	5 ON DIS OFF	6 ON DIS OFF
7 ON DIS OFF	8 ON DIS OFF	9 ON DIS OFF	10 ON DIS OFF	11 ON DIS OFF	12 ON DIS OFF
13 ON DIS OFF	14 ON DIS OFF	15 ON DIS OFF	16 ON DIS OFF	17 ON DIS OFF	18 ON DIS OFF
19 ON DIS OFF	20 ON DIS OFF	21 ON DIS OFF	22 ON DIS OFF	23 ON DIS OFF	24 ON DIS OFF
25 ON DIS OFF	26 ON DIS OFF	27 ON DIS OFF	28 ON DIS OFF	29 ON DIS OFF	30 ON DIS OFF
31 ON DIS OFF	32 ON DIS OFF	33 ON DIS OFF	34 ON DIS OFF	35 ON DIS OFF	36 ON DIS OFF
37 ON DIS OFF	38 ON DIS OFF	39 ON DIS OFF	40 ON DIS OFF	41 ON DIS OFF	42 ON DIS OFF

Scoring Direct Social Interaction Observation Form

Total:

On-task: _____total marked/ _42___total boxes *100% = ____
Disruptive: _____total marked/ __42__total boxes *100% = ____

Appendix E

Informed Consent

USU IRB Approved: Nov. 5, 2010; Page 1 of 3
Approval Terminates 11/04/2011
Protocol Number: 2734
IRB Password Protected per IRB Administrator

Department of Psychology
2810 Old Main Hill
Logan UT 84322-2810
Telephone: (435) 797-144

LETTER OF INFORMANT

Agreement of Direct Behavior Rating Scale with Systematic Direct Observation

Introduction/ Purpose . Professor Donna Gilbertson in the Department of Psychology at Utah State University with Elizabeth Popescue, a master's student candidate, is conducting a research study to find out more about how well different observation systems can be used to observe and estimate frequency of class-wide behaviors. You will fill out a brief demographic survey with basic information about yourself and then participate in your typical lab section or class and be videotaped from the back of the room. These tapes will be used by researchers to calculate percentages of behavior using three observation methods. While viewing a tape for seven minutes, an observer will record student behavior by 1) recording one student's behavior at a time every 10 seconds until all students are observed at least one time, 2) recording the behavior one of three randomly selected students every 10 seconds and 3) rating behavior of all students on a rating scale at the end of the observation. You have been asked to take part because you are attending one of the psychology classes who may choose this research as one of several offered methods to receive course credit or extra credit for participation. There will be approximately 200 total participants in this research.

Procedures If you agree to be in this research study, you will attend either your regularly scheduled lab/lecture or optional lab and be video-taped for about a one hour session in which you will be observed for on-task behavior and disruptive behavior. Classrooms will only be taped if there are twenty or more participants.

1. First, we would like you to tell us a little bit about yourself by filling out the attached brief demographic survey. Please complete this survey and turn it in to the researcher.
2. We will then ask you to sit in a seat within an area of the classroom that will be in the view of a video camera that will be set up in the back of the room during your regular lab/lecture time or an optional lab time Recording from the back of the room will decrease the chance you can be identified. This camera will be set up to tape about 20 participants during the class lesson.
3. You will be taped for about an hour during class time as planned by the instructor.
4. The video tape will be reviewed by the researchers to code observed classroom behaviors. Estimates of the percentage of classroom behaviors will be obtained

using two commonly used but resource intensive, systematic direct observation methods and compared to a simple 10 point rating scale estimate of behavior.

LETTER OF INFORMANT

Agreement of Direct Behavior Rating Scale with Systematic Direct Observation

Risks Your participation in this study is considered minimal risk. You may feel some discomfort being videotaped and there may be a small risk of a breach of confidentiality. However, measures have been taken to minimize this risk. More information is provided below under "Confidentiality."

Benefits There may not be any direct benefit to you from these procedures; however, the videos will be used to learn more about classroom observation systems that can be practically used in a school setting to accurately identify classrooms that may need support with intervention for improved class-wide behavior management systems.

Explanation & offer to answer questions Elizabeth Popescue has explained this research study to you and answered your questions. If you have other questions or research-related problems, you may reach Professor Gilbertson at (435) 797- 2034 or donna.gilbertson@usu.edu.

Voluntary nature of participation and right to withdraw without consequence Participation in research is entirely voluntary. You may refuse to participate or withdraw at any time without consequence or loss of benefits.

Confidentiality Research records will be kept confidential, consistent with federal and state regulations. Only the researchers involved in this study will have access to the data, which will be kept in a locked file cabinet in a locked office to maintain confidentiality. The researchers will view your videotaped sessions multiple times. They will be reviewed in a locked office and stored in a locked container between viewings. Further, no names will be recorded on papers or tapes to protect your privacy as much as possible. The video recordings will be destroyed after three years. If the results of this study are published, no names will be used that will reveal the identity of the participants.

Payment/Compensation You will receive course credit or extra credit for your participation in this study.

IRB Approval Statement The Institutional Review Board for the protection of human participants at USU has approved this research study. If you have any pertinent questions or concerns about your rights or a research-related injury, you may contact the IRB Administrator at (435) 797-0567 or email irb@usu.edu. If you have a concern or complaint about the research and you would like to contact someone other than the research team, you may contact the IRB Administrator to obtain information or to offer input.

Investigator Statement "I certify that the research study has been explained to the individual, by me or my research staff, and that the individual understands the nature and

purpose, the possible risks and benefits associated with taking part in this research study. Any questions that have been raised have been answered."

Donna M. Gilbertson, Ph.D.
Principal Investigator
(435) 797-2034
Donna.gilbertson@usu.edu

Elizabeth Popescue
Graduate Researcher
epopescue@gmail.com

LETTER OF INFORMANT

Agreement of Direct Behavior Rating Scale with Systematic Direct Observation

Graduate and Undergraduate student research assistants:

Evan Adams

Joanna Jenkins

Ryan Greene