

## **ABSTRACT**

WERY, JESSICA J. Interventions Employed with Students with Emotional Disturbance: A Meta-Analysis. (Under the direction of Dr. Douglas Cullinan).

The purpose of the study was to synthesize research studies that examined interventions used with students with emotional disturbance. Specific goals were to determine the extent to which interventions improve outcomes for students with emotional disturbance, whether emotional disturbance characteristics are related to the effectiveness of interventions, and identify which interventions are most effective for students with different emotional disturbance characteristics. An extensive search of the research studies identified 73 studies employing single case research design published in peer-reviewed journals between 1985 and 2010, which yielded 713 comparisons. Published single case research studies of interventions used with students with emotional disturbance indicate that interventions can change behavior and academic outcomes. Overall the effect was moderate. Students with specific ED characteristics may respond to intervention differently. Further, a disaggregation analysis indicated characteristics of ED are associated with variable effects for different intervention types. Some interventions may be more effective for students with some characteristics than others. Additional research is needed to confirm these findings.

© Copyright 2012 by Jessica J. Wery

All Rights Reserved

Interventions Employed with Students with Emotional Disturbance: A Meta-Analysis

by  
Jessica J. Wery

A dissertation submitted to the Graduate Faculty of  
North Carolina State University  
in partial fulfillment of the  
requirements for the degree of  
Doctor of Philosophy

Curriculum, Instruction and Counselor Education

Raleigh, North Carolina

2013

APPROVED BY:

---

Dr. Douglas Cullinan  
Committee Chair

---

Dr. John Nietfeld

---

Dr. Susan Osborne

---

Dr. Scott Stage

## **DEDICATION**

For all students with emotional disturbance, who struggle each and every day, may we endeavor to learn how to best meet your needs.

## **BIOGRAPHY**

Jessica J. Wery was born in Stevens Point, WI on July 20, 1973 and spent her early years growing up in Mosinee Wisconsin. She graduated from the University of Wisconsin – Milwaukee with a Bachelor of Science in Exceptional Education in 1997. Jessica received her Master of Education degree in Special Education from North Carolina State University in 2003. In between degrees, she taught at Whitewater Middle School from 1997-2000 in Whitewater, WI and after completing her Master’s degree taught for three years at West Lake Elementary School in Apex, NC. Jessica entered the doctoral program in Curriculum and Instruction at North Carolina State University in 2007, where she was also an adjunct lecturer. Her research interests include effective intervention techniques for children with emotional and behavioral disorders, and the confluence of effectiveness where research is applied in practice.

## ACKNOWLEDGMENTS

This manuscript is dedicated to all of those who have helped me along in the process completing my doctoral studies. I would like to thank my doctoral advisory committee, Douglas Cullinan, John Nietfeld, Susan Osborne, and Scott Stage for taking the time to offer their insight and encouragement while completing this program. Your efforts and perspective on my dissertation were invaluable and I appreciate every second you took to spend time aiding me in this endeavor.

I offer my sincere gratitude to my committee chair and doctoral advisor, Douglas Cullinan, who encouraged me to reach my potential by offering me opportunities that have improved my writing and research abilities. You have provided me with countless hours of mentorship and honest feedback that have made me a better scholar.

Also, I would like to thank Irene Armstrong for her responsiveness and encouragement.

Thank you to all of the teachers, professors and students who have come before me, many of who wrote articles and conducted research featured in this meta-analysis. Without their work, none of this would have been possible. In particular, I would like to thank Kimberly Vannest and John Davis of Texas A & M University, who so generously shared their knowledge and understanding of SCR effect size indices.

Finally, I would also like to give a very special thanks to my husband and daughter. Michael and Madisen, without your understanding and patience I would not have been able to complete this degree. You and your wonderful nature have made this possible.

## TABLE OF CONTENTS

LIST OF TABLES.....	vii
LIST OF FIGURES.....	ix
CHAPTER 1.....	1
Empiricism and National Interest in All Students Doing Well in School.....	2
Emotional and Behavioral Disorders.....	4
Emotional Disturbance.....	4
Outcomes of Students With ED.....	8
Using Empiricism to Address the Needs of Students with ED.....	9
Research Questions.....	9
Significance of the Study.....	10
Chapter Summary and Conclusion.....	11
CHAPTER 2.....	12
Research Questions.....	15
Forms of Research and Research Designs.....	15
Secondary Research.....	22
Extant Secondary Research.....	27
Intervention Effectiveness By Characteristic of ED.....	35
Conclusion.....	42
CHAPTER 3.....	44
Search Method.....	45
Study Inclusion and Exclusion.....	46
Coding Procedure.....	48
Data Analysis.....	56
CHAPTER 4.....	64
Studies Included.....	64
Inter-observer Agreement on Coding.....	66
Effect sizes.....	67
Summary.....	88
CHAPTER 5.....	90
To What Extent do Interventions Improve Outcomes for Students with ED?.....	91
Do ED Characteristics Moderate The Effectiveness Of Interventions?.....	95
Which Interventions Are Most Effective For Students With Different ED Characteristics?.....	96
Research Implications.....	98
Limitations.....	101
Summary.....	105
APPENDIX.....	145

Appendix A.....	146
Systematic Literature Reviews .....	146
Appendix B.....	151
Meta-analyses .....	151
Appendix C.....	160
Studies Excluded .....	160
Appendix D.....	165
Tables.....	165
Appendix E.....	239
Figures .....	239



## LIST OF TABLES

Table 1 Codification of Independent and Dependent Variables.....	165
Table 2 Included SCR Studies.....	170
Table 3 Individual Study and Comparison Effective Sizes.....	183
Table 4 Studies per Publication Year.....	207
Table 5 Included Studies Journal Publishers Distribution.....	208
Table 6 Effect Size Comparisons by Student Sex and Race.....	210
Table 7 Effect Size Comparison by Student Characteristics.....	211
Table 8 Effect Size Comparisons by Setting.....	212
Table 9 Effect Size Comparison by Research Design.....	213
Table 10 Effect Size Comparison by Primary Study Quality.....	214
Table 11 Effect Size Comparisons of Interventions Used with Students with ED.....	215
Table 12 Effect Size Comparisons of Interventions Implemented with Students with the Inability to Learn Characteristic.....	218
Table 13 Effect Size Comparisons of Interventions Implemented with Students with Relationship Problems.....	221
Table 14 Effect Size Comparison of Interventions Implemented with Students with Inappropriate Behavior.....	224
Table 15 Effect Size Comparisons of Interventions Implemented with Students with Unhappiness or Depression.....	227
Table 16 Effect Size Comparisons of Interventions Implemented with Students with Physical Symptoms or Fears.....	230

Table 17 Effect Size Comparisons of Interventions Implemented with Students having Schizophrenia-Like Features .....	233
Table 18 Effect Size Comparisons of Interventions Implemented with Students characterized as Socially Maladjusted .....	236

## LIST OF FIGURES

Figure 1	Individual Study and Comparison Effective Sizes.....	240
Figure 2	Effect Size Comparisons by Student Sex .....	242
Figure 3	Effect Size Comparisons by Student Race.....	243
Figure 4	Effect Size Comparison by Student Characteristics .....	244
Figure 5	Effect Size Comparisons by Setting.....	245
Figure 6	Effect Size Comparison by Research Design .....	246
Figure 7	Effect Size Comparison by Primary Study Quality .....	247
Figure 8	Effect Size Comparisons of Interventions Used with Students with ED .....	248
Figure 9	Effect Size Comparisons of Interventions Implemented with Students with the Inability to Learn Characteristic .....	249
Figure 10	Effect Size Comparisons of Interventions Implemented with Students with Relationship Problems .....	250
Figure 11	Effect Size Comparison of Interventions Implemented with Students with Inappropriate Behavior .....	251
Figure 12	Effect Size Comparisons of Interventions Implemented with Students with Unhappiness or Depression .....	252
Figure 13	Effect Size Comparisons of Interventions Implemented with Students with Physical Symptoms or Fears.....	253
Figure 14	Effect Size Comparisons of Interventions Implemented with Students having Schizophrenia-Like Features .....	254
Figure 15	Effect Size Comparisons of Interventions Implemented with Students characterized as Socially Maladjusted.....	255

## CHAPTER 1

The topic of emotional or behavioral disorders of children has continued to command the interest of the public and professionals. This topic is among many that has been much influenced by our culture's reliance on the scientific method as a way of knowing. For example, what we know about providing intervention for children with emotional or behavior disorders, as well as the value of what we know, is conditioned by whether or not the scientific method has been used in order to know it. Current legislative mandates also reflect and strengthen efforts of leading individuals and organizations that support evidence-based interventions with students with emotional or behavioral disorders (Odom, 2009; Odom, Brantlinger, Gersten, Horner, Thompson, & Harris 2005; Kratochwill & Stoiber, 2002).

Research and the scientific method has informed the field of emotional or behavioral disorders (EBD) of children—in clinical, counseling, community, institutional, educational, and other contexts and disciplines. In fact, the great amount of research on interventions for EBD can become an impediment to knowing—creating an inability to see the forest for the trees. Here, the scientific method again can offer ways for assembling many, sometimes disparate, research results on intervention. One of these ways, called meta-analysis, enables the user to assemble individual research findings on a particular topic in order to see the big picture, or at least a big picture.

This meta-analysis is concerned with applying a meta-analysis technique to research on interventions for students with EBD in education settings. It focuses on a particular kind

of meta-analysis technique, emotional and behavioral disorders of children, and intervention research design.

### **Empiricism and National Interest in All Students Doing Well in School**

Our culture is experiencing an increased reliance on empiricism, also called the scientific method. Evidence of this growing concept can be found in business, medicine, and the social sciences. Increasingly, decisions are made based on the analysis of measurable outcomes. Decision-making processes based on the idea that something is “believed to work” are giving way to processes that are based on measurable outcomes that are “shown to work.” This trend is also growing in education as indicated in legislation and literature (Christenson & Thurlow, 2004; ESEA, 2001; IDEIA, 2004).

Empiricism has played a role in education for a long time. It has been the foundation of special education since the early 1970s, and empiricism is increasingly prevalent in general education policy. Federal and state mandates reflect this increasing emphasis on empiricism (Christenson & Thurlow, 2004). The U.S. Department of Education has called for greater rigor in educational research (IES, 2010) and for the use of evidence-based practices in schools. Recent federal laws including Elementary and Secondary Education Act of 2001 (“No Child Left Behind”) and the Education Sciences Reform Act of 2002 reflect the trend toward accountability, objectively measured outcomes, evidence-based practices, and scientific rigor. The most recent reauthorization of the federal special education law, Individuals with Disabilities Education Improvement Act of 2004 (IDEIA), also addresses

empirical and scientifically-grounded practices for students with special education needs (National Council on Disability, 2004).

These mandates and laws both reflect and strengthen efforts of leading individuals and organizations that support evidence-based interventions with students, especially special education students and children with emotional or behavioral disorders (Odom, 2009; Odom et al., 2005; Kratochwill & Stoiber, 2002). Further, the mission statements of more and more professional journals emphasize the publication of scientific research reports (Clarke, Dunlap, & Peck Stichter, 2002).

What constitutes scientific evidence-based or empirically-based interventions? These and related terms have been easier to speak of than to define. Several individuals and organizations have provided thoughtful discussions of these points. The U.S. Department of Education's Institute of Education Sciences (IES) has developed quality indicators for research to judge empirical evidence for educational practices (U.S. Department of Education, 2002). In addition, the IES manages a "What Works Clearinghouse" (WWC) as a repository of scientific evidence for educational practices (Institute of Education Sciences, 2011) and a dissemination source for results of its evaluations of educational practices. The WWC aim is to improve students' outcomes by increasing educator access to empirically based practices. The Coalition for Evidence Based Policy is another organization that establishes research standards, uses them to evaluate intervention practices in various disciplines, and creates databases of intervention programs that meet stringent research evidence of effectiveness (Top Tier Evidence, 2011; PL. 110-161, PL. 111-8).

## **Emotional and Behavioral Disorders**

More than 20% of U.S. children and adolescents currently have, or previously have had, a significant EBD (Jaffee, Harrington, Cohen, & Moffitt, 2005; Merikangas, He, Brody, Fisher, Bourdon, & Koretz, 2010; National Research Council and Institute of Medicine, 2009; U.S. Department of Health and Human Services, 1999; Walker, Nishioka, Zeller, Severson, & Feil, 2000). EBD is an unofficial, nonspecific, broad category that includes children and adolescents with a diagnosed mental disorder, students identified for the emotional disturbance (ED) category of special education, and young people who have not been officially identified but whose social and personal maladaptation, if brought to professional attention, would readily justify their identification. My present meta-analytic study focuses on students identified in the emotional disturbance category.

## **Emotional Disturbance**

The federal special education law (PL 94-142) abbreviated as IDEIA was signed into law in 1975 and reauthorized most recently in 2004. It specifies 13 kinds of education disability, including ED. Students in any of the disability categories may exhibit emotional or behavior problems, but ED is the category of education disability set aside for students whose emotional and behavior problems are the most prominent aspect of their education disability. Students with ED are those who are certified to have met the “federal definition”, that is, the definition of ED stated in IDEIA regulations.

At the most recent accounting, about 442,000 students were identified as having ED (U.S. Department of Education, 2009). This is less than one percent of public school students and about eight percent of all students with disabilities.

**Federal definition of ED.** The federal definition of ED is as follows:

(i) The term means a condition exhibiting one or more of the following characteristics over a long period of time and to a marked degree that adversely affects a child's educational performance:

(A) An inability to learn that cannot be explained by intellectual, sensory, or health factors.

(B) An inability to build or maintain satisfactory interpersonal relationships with peers and teachers.

(C) Inappropriate types of behavior or feelings under normal circumstances.

(D) A general pervasive mood of unhappiness or depression.

(E) A tendency to develop physical symptoms or fears associated with personal or school problems.

(ii) The term includes schizophrenia. The term does not apply to children who are socially maladjusted, unless it is determined that they have an emotional disturbance. (*Federal Register*, 2006, p. 46756)

The first section of this definition contains key points about the distinguishing characteristics of the disability and the intensity and duration of those characteristics. In order



for a student to qualify as having ED, he must exhibit at least one of the five characteristics (A—E) in an enduring, intensive, significant way that adversely affects his educational performance. The federal definition also names two other conditions that do (schizophrenia) or may (socially maladjusted) qualify as ED.

One reason the federal definition of ED is important is that it affects, or at least could affect, funding of special education. In order to qualify for a federal per capita stipend to help defray the excess cost of educating a student with a disability, each state's education agency must attest that any student identified with one of the 13 disabilities meets the federal definition of that disability (Palmaffy, 2001). Specifically, for a student to qualify as a student with ED, the state annually must assure the U.S. Department of Education that the student meets the federal ED definition. This is a compelling justification for using the federal definition in the present meta-analysis to determine which students should be considered as ED in the search for effective interventions.

**State terms and definitions.** As an aside, states may and do use terms other than “emotional disturbance” for this category of student. Each state has authorized terms and definitions for various educational disabilities. These terms and definitions can be identical to the federal ones, or can be different. In either case the state must assure that children in a state-defined category of disability also qualify under a federal category.

It has been argued that variations across states' definitions of ED make it difficult to compare students identified in one state to students in another state (Gresham, 2007; Hallahan, Keller, & Ball, 1986), or that a student who formerly qualified for ED in one state

may no longer qualify in another state to which he moved. This argument has always been abstract; no real cases of this occurring have been presented. Nonetheless, its merit is probably greatly diminishing, as states' definitions of ED have become increasingly similar to each other as they have become more similar or identical to the federal definition over time (Wery & Cullinan, 2011). Further, the terminology used to describe ED is increasingly consistent. Nearly half of all states also use the federal term "emotional disturbance". However, some states do use a variety of a combination of the following terms: (a) behavior, (b) emotional, or (c) social in conjunction with one of the following: (d) disability, (e) disorder, or (f) impairment. Crossing the first three adjectives with the second three nouns yields terms such as behavior disorder, behavioral disability, emotional and behavioral disorder, emotional disability, or emotional impairment (Wery & Cullinan, 2011).

**Criticism of the federal definition.** The federal definition of ED has been a target of professional criticism. For example, it has been faulted as having philosophical deficiencies, lacking practical applications, containing vague concepts, and being illogical, self-contradictory, and redundant (Bower, 1982; Kauffman, Mock, & Simpson, 2007; Nelson, Rutherford, Center, & Walker, 1991; Skiba & Grizzle, 1991; Tankersley, Landrum, & Cook, 2004). Professional and advocacy groups (e.g., Council for Children with Behavioral Disorders, National Mental Health and Special Education Coalition) have proposed alternative terms and definitions (e.g., Forness & Knitzer, 1992) to the U.S. Congress and Department of Education. However, these attempts to modify the federal definition of ED were unsuccessful, with the exception of minor changes in wording (e.g., from an earlier

term, “serious emotional disturbance”, removal of the modifier “serious”; Federal Register, 1993). The federal definition of ED has remained essentially stable since 1975.

### **Outcomes of Students With ED**

Students with ED have exceptionally poor educational, behavioral and social outcomes (Bradley, Doolittle, & Bartolotta, 2008; Kauffman 2005; U.S. Department of Education, 2000). This unfortunate fact gives greater urgency to efforts to identify powerful school based interventions aimed at improving academic and behavioral outcomes for students with ED.

**Academic outcomes.** On average, students with ED exhibit significant academic deficits (Wagner, Kutash, Duchnowski, Epstein, & Sumi, 2005). They experience academic difficulties such as low or failing grades, many course failures, high grade retention rates, and an increased likelihood of dropping out of school (Reschley & Christenson, 2006). No other category of educational disability has a high school graduation rate as low as ED (below 40%; U.S. Department of Education, 2009).

**Behavioral outcomes.** As a group, students with ED have serious, multiple, and complex problems that begin early in life. They tend to have multiple emotional and behavioral problems (Cullinan & Epstein, 2001) rather than just one. Almost half of all students with ED (47.7%) have been suspended or expelled at least once (Wagner, Kutash, Duchnowski, Epstein, & Sumi, 2005). They ultimately have high rates of unemployment, involvement with the criminal justice system, and substance misuse (Kauffman & Landrum, 2009; Wagner, Newman, Cameto, Garza, & Levine, 2005).

## **Using Empiricism to Address the Needs of Students with ED**

To the extent empiricism has increased in influence in the evaluation of interventions for students with ED, there remains a great deal more that needs to be learned about optimal interventions for this population. Further, there remains a large gap between what is known about effective interventions and what practitioners implement as classroom and school interventions for students with ED.

Historically, substantial numbers of teachers have been unaware of research or confused by intervention research results (Wehby, Symons, Canale, & Go, 1998). Conflicting reports about intervention research results have further frustrated teachers, which may explain why empirically supported interventions are not routinely implemented in the classroom (Carnine, 2000). Problems related to communicating research results to teachers, school and district administrators, and others involved in assuring appropriate education to students with ED are not likely caused by a single impediment to communication, and will not be solved readily. However, better summaries of existing research on interventions for ED will be helpful. Summarizing intervention research results via meta-analytic techniques can help remedy this problem. This is a main thrust of the present dissertation.

### **Research Questions**

As I describe in Chapter 2, meta-analysis summarizes a number of data-based studies by assigning a value to each relevant comparison in a particular study, then combine the assigned values across all studies (Wilson & Lipsey, 2001). This produces a quantitative estimate of the strength of the intervention in the summarized studies. This estimate of

strength is often called “effect size.” Assuming the researcher has identified all or nearly all appropriate studies bearing on that intervention, effect size can be an accurate and readily communicated indicator of the intervention’s strength. It is possible to compare effect sizes for different interventions.

My general research goal was to identify and examine was to examine relevant ED intervention studies on intervention for students with ED in order to determine the effectiveness of such. I assigned one or more effect size statistics to each study, then aggregated and disaggregated these values in order to address my research questions. In addition, I analyzed these studies to address the possibility that the characteristics of ED, as found in the federal definition, moderate the effectiveness of interventions for ED.

The goal of my study is to answer the following research questions:

Do interventions improve outcomes for students with ED?

Are ED characteristics related to the effectiveness of interventions?

Which interventions are most effective for students with different ED characteristics?

Answers to these questions may be useful to improve ED students’ outcomes by identifying which interventions are most effective for students manifesting various ED characteristics. Answers may simplify communication of such information to practitioners serving students with ED.

### **Significance of the Study**

This study aggregated and synthesized empirical evidence found within a body of relevant published studies on interventions used with students with ED. Results identified

information about interventions that might provide a clearer picture to education practitioners and researchers about which interventions have more empirical support and are most likely to improve outcomes for students with ED. This study will add to our field's knowledge of the overall value of different interventions for students with ED. Additionally it will contribute examples of how employ a method of aggregating single-case research studies.

### **Chapter Summary and Conclusion**

Legal mandates emphasize the need for the identification and use of empirically supported interventions for students receiving special education services, including those with ED. These students have specific characteristics and related needs, which can be mitigated with the careful application of appropriate interventions. This study will add to the current research on the effectiveness of interventions used with students with ED.

In Chapter 2 of this manuscript, I review the foundation of secondary research, including extant secondary research on interventions used with students with ED. Chapter 3 describes the research design and methodology used in this study. In Chapter 4, I present the results; in Chapter 5, I discuss those findings and detail their implications.

## CHAPTER 2

As stated in Chapter 1, students with emotional disturbance (ED) exhibit at least one of five characteristics specified in the federal definition of ED. The five characteristics can be briefly termed (A) Inability To Learn, (B) Relationship Problems, (C) Inappropriate Behavior, (D) Unhappiness or Depression, and (E) Physical Symptoms or Fears. The characteristic(s) must be evident in an enduring, intensive, significant way that adversely affects the student's educational performance. Of course, student's functioning in personal, social, and other domains is likely to be affected as well.

Over time, many interventions have been created and implemented to improve schooling for these challenging students. There is quite a bit of research on these interventions, as well as some research reviews on the topic, but we need to know more about relative effectiveness of interventions for ED. While the field of special education has progressed, there is much room for continued improvement in understanding interventions for students with ED.

### **Problems for Research Reviews**

**Problems with the nature of participants.** Inconsistent and varying definitions and terms have been used to describe students with ED and with emotional or behavior disorders (EBD). As noted in Chapter 1, EBD is a broad, nonspecific descriptor of a population of children of which a smaller part consists students with ED. However, this discrimination is not consistently followed. Some researchers seem to assume that if a student exhibits EBD, that student may as well be identified with ED. Other researchers use EBD or another

nonspecific term to mean ED (that is, meeting the federal definition). This inconsistent term use found in the research corpus impedes the interpretation of the research base by both research reviewers and practitioners.

**Variations among participants.** Studies also differ in their participants' age, grade, sex, special education placement environment, and other variables. Some of these differences may affect the value of interventions. Further, some studies do not clearly report their participants' demographic information, again making it difficult to generalize the findings in everyday practice.

**Dependent variables.** Studies also use different dependent or outcome variables to gauge the merit of an intervention. Some measures are more fine grained or sensitive and result in outcome changes that may appear to be larger than those that use measures that are more granular. Some dependent variables may be measured with specific behaviors, while others measure larger phenomenon. The former is likely to show greater change in behavior than the latter. For example, behavior counts of a specific behavior (e.g., hand raising) may be more sensitive to change than improved externalizing behavior measured by an global behavior rating scale.

**Research designs.** As described later in this chapter, there are two major forms of research design used to study interventions for students with ED. They are often called group comparison designs and single-case research (SCR) designs, and they will be described later in this chapter. Both kinds of design, if implemented properly, can yield valuable evidence about the merit of educational interventions, and both are widely used in individual studies of



interventions for students with ED. In reviewing interventions, however, it is problematic to combine results of group comparison and SCR designs (Lipsey, & Wilson, 2001).

Similarly, the selection of a SCR design may also impact the outcome and perceived value of an intervention. While no evidence could be found to support this notion, some may wonder if some research designs may more sensitive to outcome change, inflating the appearance of true behavior change compared to those designs that may not be as sensitive. Given equally powerful interventions, are some designs more sensitive to behavior change than others?

Although the aforementioned and other obstacles plague those who attempt to review research, intervention research reviews are critical to continued progress. We may know more than we think we know about some aspects of intervention for students with ED, and a research review is a major way through which we can arrive at this revelation. This meta-analysis may point to previously unrecognized reasons that interventions appear not to work consistently, limitations on the conditions under which they do work, or kinds of outcomes that should and should not be expected. This meta-analysis comes to grips with those and other obstacles, and summarizes the extant SCR studies on interventions used with students with ED. The purposes are to identify those interventions that have evidence of effectiveness, and whether or not the characteristics of ED moderate the effectiveness of those interventions.

## **Research Questions**

The present meta-analytic study aggregates data from SCR on interventions used with students with ED. This study addresses the following questions, which constitute the goals of my research:

Do interventions improve outcomes for students with ED?

Are ED characteristics related to the effectiveness of interventions?

Which interventions are most effective for students with different ED characteristics?

The aggregation of data from a pool of studies can form a clearer picture of the effectiveness of interventions used with students with ED. This information could be used to improve these students' outcomes by identifying which interventions are most effective for students manifesting ED characteristics.

## **Forms of Research and Research Designs**

As preparation for the presentation of my research method in Chapter 3, this chapter discusses two levels of research, primary research and secondary research, as well as two very general types of research design, group comparison design and SCR design. This discussion also sets the stage for my summary of existing secondary research reviews on the effectiveness of interventions used with students with ED, which makes up the balance of the present chapter.

**Primary research.** Primary research is the performance and report on a single empirical research study (Cooper, 2005). Primary intervention research is research that is directly conducted by a researcher who (a) systematically manipulates one or more

independent variables (e.g., specific interventions) as applied to participants (e.g., students with ED), and (b) measures changes in one or more dependent variables (e.g., behavior, emotion, or academic outcomes). Primary intervention research study results are often reported in peer-reviewed journals but may be presented elsewhere as well. Primary research studies take advantage of either a group comparison design or a SCR design.

***Group comparison design research.*** In a primary intervention research study using group comparison design, there are two or more groups of participants. If two groups are employed, the researcher manipulates an independent variable by applying it to one group (intervention group) but not the other (control group). If multiple groups are employed, the researcher applies variations of the intervention to different groups while still withholding it from the control group. Thereafter, all participants are measured on the dependent variable(s), results are summarized within each group, then compared across groups (especially intervention group[s] to control group). Differences between the control and intervention(s) groups on the dependent variables are attributed to the operation of the intervention, as long as the group difference is large enough and various “threats to internal validity” (Campbell & Stanley, 1966; Gersten, Fuchs, Compton, Coyne, Greenwood, & Innocenti, 2005) have been controlled. In brief, group comparison research seeks to identify between-group differences.

Data collected through group comparison design research can be analyzed with parametric statistics. Hypothesis testing, in which statistical significance is determined, can be conducted by calculating p-values. Hypothesis testing and p-values are no longer

considered sufficient in group comparison reporting (APA, 2009). They should be supplemented by measures of “magnitude of effect” and an indication of reliability, or trustworthiness of the results. Together, effect size (ES) and confidence intervals (CIs) meet this need.

ES can be interpreted as the amount of variance in the dependent variable that is accounted for by an intervention (Trusty, Thompson, & Petrocelli, 2004). CI is a range of values that is likely, at a specified level of confidence (usually 95%), to contain the ES. For example, suppose an ES of 0.62 has a CI of 0.73 — 0.54; then, we can be confident 95% of the time that the true ES falls between 0.54 and 0.73.

While there are a number of methods for calculating ES, a simple and common one is called standardized mean difference. It is obtained by dividing the difference between the control group mean and the intervention group mean by the pooled standard deviation of both groups. This expresses the group difference for a single comparison in terms of the standard deviation. This ES is equivalent to a z-score, so an ES of 1.0 represents one standard deviation from the mean in the standard normal distribution (Jensen, Clark, Kircher, & Kristijansson, 2007). Together, ES and CI allow readers to evaluate the stability of results across group comparison research samples, designs, and analyses (Wilkinson & Task Force on Statistical Inference, 1999). Commonly, an ES of 0.80 or higher is interpreted to be a large effect, 0.50—0.79 to be a medium effect, and 0.20—0.49 to be small effect (Cohen, 1988).

Because it is a standardized difference between means, ES values of different studies can be compared to one another and otherwise manipulated. A main use of ES in group comparison studies is to permit comparisons from diverse studies to be combined in various ways, thus providing an empirical approach to summarizing results in certain secondary research reviews of group comparison studies. As noted, my dissertation study is concerned not with group comparison studies but with SCR studies. I present this information on ES of group comparison research, as a contrast with SCR studies, for which there is no similarly straightforward way to compute an ES. As I relate below, there have been several attempts to develop something equivalent to ES to serve as value to summarize comparisons made in SCR design studies.

*SCR design research.* While it is widely agreed that evidence obtained through group comparison research can provide the empirical evidence on interventions (National Research Council, 2002), much of the intervention research conducted in special education is not group comparison research. Instead, SCR is more widely used with the relatively small populations of students with special education needs. SCR is experimental research that is consistent with the scientific method (Kazdin, 2011). The corpus of SCR studies contains valuable information about the effectiveness of interventions used with students with disabilities, including ED.

The premise of SCR is that a specific dependent variable (e.g., behavior, emotion, or learning outcome) is measured repeatedly both in the absence and in the presence of a specified intervention (independent variable; Tawney & Gast, 1984). Data from the repeated

measurement of the dependent variable is graphed, with level of dependent variable on the Y-axis, and recording sessions on the X-axis. If the dependent variable is consistently better while the intervention is present and worse while the intervention is absent, this constitutes evidence that the intervention produced that improvement in the outcome variable (Gast, 2010). There are several SCR designs (e.g., withdrawal, multiple baseline) that are widely used in research on interventions for students with ED. They are described in Chapter 3, as part of the study method.

*SCR and ES.* SCR traditionally has relied on visual analysis of the graphed outcome data to detect large, clinically important results (Parsonson & Baer, 1992), rather than on significance tests of intervention-absent versus intervention-present contrasts. This position is logical and defensible, but visual analysis is not foolproof. Visual analysis is often insufficient to discriminate smaller effects, which can sometimes be important (Glass, 1997). The apparent magnitude of an effect can be manipulated by, for instance, changing the scale of the Y-axis. Also, studies have shown that visual analysts disagree about the extent of intervention effect in a considerable proportion of cases (Matyas & Greenwood, 1990).

However, consistent with the aforementioned trend toward greater empiricism in research in education, many have called for an empirical measure of SCR effects (Scruggs & Mastropieri, 1998). An early proposal for ES in SCR (Scruggs & Casto, 1987) was the approach called percent of non-overlapping data (PND). It remains the most commonly used form of ES for SCR studies. PND is described in a later section of the present chapter, along

with an improvement on PND that I use in reviewing intervention studies for students with ED.

Parenthetically, because SCR data generally do not meet typical parametric assumptions of normality, homogeneity of variance, and independent errors, parametric statistics suited for group comparison research are not appropriate for use in SCR. Parker and colleagues (Parker, Vannest, & Davis, 2011) proposed the term “effect index” to denote a measure of effects in SCR. Effect index is characteristically different from the parametric statistical ESs derived from group comparison studies. Still, Parker et al (2011) contend effect indices, although non-parametric, are like ESs in that they provide an objective measure that can be used to compare results across SCR studies. However, because the term effect index has not yet become commonly used, here ES is used for both parametric and non-parametric effect measures.

In SCR, measures of ES offer several advantages over visual analysis alone. When used with strong, internally valid research designs, ES provide an objective measure of intervention strength, a more precise and dependable measurement and an efficient method for comparing results across SCR studies (Carver, 1978; Mitchell & Hartmann, 1981; Parker & Hagan-Burke, 2007; Rosnow & Rosenthal, 1989; see also Durlak, 2002; Kirk, 1996; Rosnow & Rosenthal, 1989; Wilkinson & Task Force on Statistical Inference, 1999).

It is worth emphasizing that, for both group comparison research and SCR, causality is determined by the rigor of the experimental design, and not by the ES statistic (Gersten, Fuchs, Compton, Coyne, Greenwood, & Innocenti, 2005). Therefore, ES must be interpreted

within the context of the quality of the experimental design. In the case of a flawed design, a large ES does not indicate that the intervention was successful (Parker & Brossart, 2003; Parker et al., 2005). Several sources have offered guidelines for evaluating the rigor of experimental designs, such as Division 12 of the American Psychological Association (APA; Lonigan, Elbert, & Johnson, 1998) and Division for Research of Council for Exceptional Children (CEC; Horner, Carr, Halle, McGee, Odom, & Wolery, 2005). The Institute of Education Sciences published SCR design standards (WWC, 2010) that have been included in the present meta-analysis.

The application of an ES method yields a number that is meant to summarize the effect of an intervention. The method can be consistently applied and used by different researchers and across multiple studies. This brings up this researcher's main interest in ES statistics for SCR studies: their suitability for helping summarize diverse comparisons in a meta-analysis. Further, some ES methods allow for the calculation of CIs, which provide a gauge of the precision, dependability and reliability of results (Parker et al, 2011). As noted earlier, CI describes the range within which the true ES lies with a selected degree of confidence (typically, 95% confidence is selected). Thus CIs can be used, similar to p-values, to describe the extent that a similar effect can be obtained by chance alone (Wilkinson & Task Force on Statistical Inference, 1999; APA, 2001). For example, one could compare the ESs of two SCR intervention comparisons by observing their respective 95% CIs: if the CIs do not overlap, we can be 95% confident that their true values are different.



In summary, ESs summarize intervention effects in numerical form. Therefore they allow for the efficient comparison of intervention effects, and in particular, allow the aggregation of effects across studies. This last point is an advantage of ES computation over visual analysis of SCR. As the body of primary SCR studies has increased, researchers have needed to develop methods to integrate the results in that body of work. Secondary research methods were developed to integrate and aggregate the existing primary research. In primary research, participants (e.g., students with ED) are the subjects of the analysis; in secondary research, the ESs derived from the primary research are the subjects of analysis. Secondary research thereby provides a composite picture of what is known via primary research.

### **Secondary Research**

Secondary research involves locating, summarizing, and synthesizing existing primary research studies (or other literature). Until several decades ago nearly all secondary research in the social sciences was qualitative review of a particular topic (also called narrative review or selective literature review). Later, more scientific methods of research study review came into prominence, including the systematic literature review and the meta-analysis.

**Narrative review.** In narrative literature reviews of primary research studies, those primary studies are selected in a somewhat unsystematic and selective manner. Often these reviews are verbal summaries in chronological order of study publication (Walberg, 1985). This form of review can be problematic unless the reviewer has knowledge of and access to all appropriate instances of the research to be reviewed. Also, in this kind of review it is

difficult to know what were the reviewer's criteria for including and excluding studies, making the narrative research review harder to replicate and thus less scientific.

**Systematic literature review.** An advance in research literature reviews is the systematic literature review. In a systematic review, the reviewer clearly states key aspects of locating and evaluating studies for the review, and reporting results and conclusions based on that review (Cooper, 2005). For example, the reviewer clearly states search strategies, inclusion and exclusion criteria, and criteria for subdividing studies according to demographic, study feature, and other variables (e.g., age of participants, kind of dependent variable or research design). Thus readers know the reviewer's method for the review, and can even replicate it as appropriate. While other definitions of systematic literature review exist, this definition is used here.

**Meta-analysis.** A third kind of research review is a meta-analysis. A meta-analysis conforms to the requirements of the systematic literature review. In addition, there is a quantitative summary of the extent to which application of the intervention is associated with a difference in the dependent variable(s). Often this is done through the calculation of at least one ES statistic for each study included, then combining the effect sizes across all the relevant studies. Other ways to summarize effects from individual comparisons are available (Parker, Vannest, & Davis, 2011).

In a meta-analysis, empirical procedures include explicit and transparent search procedures and primary study inclusion criteria set *a priori*. For example, the reviewer reports which databases, journals and date ranges were searched, and specifies study

characteristics to be examined before beginning her or his analysis. Moreover, detailed objective coding schemes are created from which features and outcomes of each study are extracted and quantified. Often the reliability of the coding process is measured through inter-rater reliability testing. Finally, statistical methods are used to quantify treatment outcomes overall, and as may be contingent on moderating variables across the primary studies. These quantitative methods of meta-analysis have been well described (e.g., Glass, McGaw, & Smith, 1981; Lipsey & Wilson, 2001; Rosenthal, 1991; Wolf, 1986).

As the corpus of research in special education has grown, researchers have used meta-analytic techniques to synthesize cumulative group design research findings in special education (Forness & Kavale, 1994). The majority of this meta-analytic research in special education has been conducted on group design research, as originally proposed by Glass. However, as discussed earlier, most intervention research in special education uses SCR designs. This has created a demand for methods to create ES statistics appropriate for the aggregation of SCR data in meta-analyses.

**Meta-analysis of SCR.** Just as a variety of statistical methods can be used to aggregate group-design primary research, statistical methods have been developed to aggregate data from SCR. No one meta-analytic approach to SCR is without limitations (Parker, Vannest, & Davis 2011), and there is currently no consensus on one best approach. Several different general approaches have been described, including “non-overlap” methods.

***Non-overlap methods.*** There are several non-overlap methods available for creating a numerical ES for SCR phase comparisons, thus permitting the possibility of aggregating ESs

from multiple SCR studies. Among these possible methods are extended celeration line (White & Haring, 1980); percent non-overlapping data (PND; Scruggs & Casto, 1987); percent exceeding the median (Ma, 2006); percent of all non-overlapping data (PAND; Parker, Hagan-Burke, & Vannest, 2007); improvement rate difference, nonoverlap of all pairs (Parker & Vannest, 2009); and tau for nonoverlap with baseline trend control (Parker et al., 2011). All non-overlap methods provide an indication of performance difference between baseline (control) and intervention (experimental) phases of the SCR study; all non-overlap methods are sensitive only to data overlap, not to distance (i.e., phase differences in level or magnitude). Of the non-overlap methods, PND has been the most commonly used; however, it has been also widely criticized.

*Percent of non-overlapping data.* In the method of PND, one determines the percentage of data in Phase B that exceed (in the desired direction of behavior change) the single most extreme data point in Phase A (Scruggs & Casto, 1987). This percentage may range from 0 to 100 percent non-overlap. In the PND method there is no assumption or concern with any existing baseline trend. Like many ES indices in SCR, this index does not relate to other commonly used metrics, but its authors have offered a general interpretation guideline:  $PND > 70$  percent for effective interventions,  $50 \text{ percent} < PND < 70$  percent for interventions of questionable effectiveness, and  $PND < 50$  percent for interventions with no effect (Scruggs & Mastropieri, 1994).

PND has the advantages of being easy to calculate and well correlated with visual analysis. Unfortunately, PND showed poor ability to discriminate among approximately one

third of the data sets in a test of it. PND shows pronounced floor and ceiling effects, failing to distinguish among the lowest 12% or highest 18% of data sets (Parker, et al., 2011). Further, PND has no known sampling distribution; this prevents the calculation of CI, with the result that PND can provide no information about the likelihood that results are due to chance (Parker, Hagan-Burke, & Vannest, 2007).

*Percent of all non-overlapping data.* Percent of all non-overlapping data (PAND) is a nonparametric effect size that is based on Cohen's (1988) description of the relationship between effect sizes and percentage of data that do not overlap. PAND is described as the percent of all data remaining after removing the minimum number of data points that eliminate all data overlap between phases A and B (Parker, Hagan-Burke, & Vannest, 2007).

Like PND, PAND does not control for problems associated with trends in the data of baseline or intervention phases. However, PAND is free of many of the difficulties associated with PND. Most importantly, PAND can be converted to Phi coefficient, a commonly accepted statistic with which a CI can be computed (Burns, Coddling, Boice, & Lukito, 2010). CIs can be used to identify statistically significant difference between comparisons. When Phi is used with PAND, it has more statistical power than PND. Cohen (1988), suggests interpreting Phi coefficients less than 0.19 as negligible, 0.20-.49 as small, 0.50-.79 as moderate, and  $>0.80$  as strong.

As described in Chapter 3, I use “Robust Phi” (hereafter referred to as *RPhi*) in this meta-analysis. RPhi is Phi with balanced marginals in its 2-x-2 chi-square table. This makes RPhi more stable than Phi values from unbalanced matrices (Liu, 1980). RPhi, a

nonparametric ES, is not limited by parametric assumptions. It has a known sampling distribution, which also allows for the calculation of p-values and CIs. RPhi can easily be computed and translated into a readily interpretable effect size metric (Schneider, Goldstein, & Parker, 2010). Given these attributes, I decided to use PAND with RPhi in my meta-analysis to compare phases in the qualifying SCR studies.

### **Extant Secondary Research**

Researchers and professional organizations have attempted to review and find trends in existing research on interventions used with students with ED and EBD. Lakin (1983), and Skiba and Casey (1985) conducted the seminal secondary research on interventions used with students with ED. Lakin (1983) found that the majority of studies reviewed selected participants by their presence in a setting without any attempt to substantiate, quantify, or qualify the cases of those nominations. A couple of years later, Skiba and Casey (1985) attempted to identify which treatments and interventions were most effective for students with ED, using a form of meta-analysis. They concluded that interventions targeting classroom behavior and those attributable to a behavioral orientation were powerful and robust. However, the authors pointed out that the primary research was fraught with definition and reporting problems, limiting the utility of results.

In the early 1990s, experts in the field formed an informal working group meet in Peacock Hill, Charlottesville, Virginia to identify successful practices and programs used with students with ED. While they did not perform an exhaustive literature review, these authorities identified several successful strategies for teaching this challenging population

including: use of systematic, data-based interventions; continuous assessment and monitoring of progress; providing for new skill practice; matching treatment to the problem; utilizing multi-component treatments; programming for transfer and maintenance; and commitment to sustained intervention (The Peacock Hill Working Group, 1991).

Since then many researchers have used systematic literature review and meta-analysis practices to aggregate the effects of interventions used with students with ED and with EBD. I conducted a systematic search using ERIC and Psych Info to identify the secondary research on interventions used with students with EBD or ED, in schools, published in peer-reviewed journals. This search revealed 10 systematic literature reviews, one review of literature reviews, and 18 meta-analyses. Each secondary research effort aimed to answer a unique set of research questions, and consequently each varied not only by methodological approach, but also by the primary study inclusion criteria.

**Types of secondary research.** Systematic literature reviews and meta-analytic approaches have been used most often. However, others have used other synthesis approaches including: “review of reviews” and “best evidence synthesis approach.”

***Systematic literature reviews.*** Literature reviews summarizing the extant literature on the effectiveness of interventions used with students with behavioral and emotional problems have been conducted by Coleman and Vaughn (2000) (teacher perceptions of reading instruction); Evans, Harden, and Thomas (2004) (effective interventions for mainstream settings); Gulchack and Lopes (2007) (interventions used internationally); Heckaman, Conroy, Fox, and Chait (2000) (functional assessment-based interventions); Lewis, Hudson,

Richter, and Johnson (2004) (scientifically supported practices); McIntosh, Vaughn, and Zargoza (1991) (social interventions); Reddy, DeThomas, Newman, and Chun (2009) (school-based prevention and intervention programs); Scruggs, Mastropieri, and Richter (1985) (peer tutoring); as well as Spencer (2006) (peer tutoring). See Appendix A for summaries of each.

***Meta-analyses.*** Meta-analyses on this topic have been conducted by Ang and Hughes (2001) (social skills training); Beelman, Pflingsten, and Losel (1994) (training social competence); Benner, Nelson, Ralston, and Mooney (2010) (reading instruction on the reading skills); Durlak, Furhman, and Lampman (1991) (cognitive-behavior therapy); Gage, Lewis, & Stichter (2012) (functional behavioral assessment-based interventions); Ghafoori and Tracz (2001) (cognitive-behavioral therapy); Griffith, Hurley, and Hagaman, (2009) (treatment integrity of literacy interventions), Joseph and Eveleigh (2011) (self-monitoring on the reading performance); Losel and Beelman, (2003) (skills training on preventing antisocial behavior); Mathur, Kavale, Quinn, Forness, and Rutherford (1998) (social skills interventions); Mooney, Ryan, Uhing, Reid, and Epstein (2005) (self-management interventions targeting academic outcomes); Nelson, Smith, Young, and Dodd (1991) (self-management); Pierce, Reid, and Epstein (2004) (teacher-mediated interventions on academic outcomes); Quinn, Kavale, Mathur, Rutherford, and Forness (1999) (social skill interventions); Ryan, Reid, and Epstein (2004) (peer-mediated intervention); Schneider and Byrne (1985) (predictors of successful transition); Spencer (2006) (peer tutoring), as well as



Templeton, Neel, and Blood (2008) (math interventions). See Appendix B for summaries of the above.

***Review of secondary research.*** A handful of secondary research studies do not fit neatly in the categories above. One example is Maag (2006). Maag reviewed 13 literature reviews and meta-analyses of social skills training used with students with behavioral and emotional problems, ED, or EBD. He reported evidence that supports the use of selection of socially valid behaviors, programming for generalization, and including peer groups. In addition, he noted that social skill training should be tailored to individual student need.

Nelson, Lane, Benner, and Kim (2011) conducted an analysis that they termed a “best evidence synthesis approach” which is described as a combination of the quantification of effect sizes and the systematic study selection with attention to the methodological and substantive issues of typical narrative reviews. The meta-analytic approach utilized Hedges’  $g$  (Hedges, 1984) to aggregate data from four group or quasi-experimental studies on effects of literacy instruction on social adjustment. The primary studies included students with or at-risk for reading and/or behavior problems. Social adjustment outcomes from direct observation of outcomes were found to be small to moderate improvements in behavioral outcomes, yet teacher reported measures revealed no positive effects or negative effects. The results of their synthesis suggest that effective literacy instruction does not appear to have collateral effects on social adjustment of children.

**Varied primary inclusion criteria.** It is important to be aware of the similarities and differences among secondary research reports when interpreting and comparing the

outcomes. Secondary research reports should include a clear description of the search method used to identify the primary studies as well as the inclusion criteria. Of course, these are among the requirements of good systematic literature reviews and meta-analyses as well.

**Quality.** Some secondary research included only primary research reports that met a set of quality or rigor guidelines. Some included only those reports that had been peer-reviewed, arguing that the peer-review process distinguishes the more rigorous from the flawed. Still others did not dismiss research because of technical or design flaws, and included both published and unpublished studies.

**Study source.** Secondary research may also vary by the location of the research report. Some included only those research reports published in peer-reviewed journals. Others included dissertations and other non peer-reviewed sources (e.g., conference proceedings). Some restricted their inclusion criteria by publication date, and others included any studies available.

**Research method.** Secondary research has also varied by the research method of primary studies included in the synthesis. Of those that utilized meta-analysis, three included only group experimental comparison studies (Durlak, Furhman, & Lampman, 1991; Quinn, Kavale, Mathur, Rutherford, & Forness, 1999; Schneider & Byrne, 1985.) Four meta-analyses included experimental and quasi-experimental group comparison studies (Ang & Hughes, 2001; Beelman, Pfingsten & Losel, 1994; Hoagwood, Serene Olin, Kerker, Kratchowill, Crowe, & Saka, 2007; Nelson, Lane, Benner, & Kim, 2011). Four meta-analyses included both experimental group and SCR studies (Benner, Nelson, Ralston, &

Mooney, 2010; Coleman & Vaughn, 2000; Olmeda & Kauffman, 2003; Pierce, Reid, & Epstein, 2004). Three meta-analyses included only SCR (Gage, Lewis, & Stichter, 2012); Mathur, Kavale, Quin, Forness, & Rutherford, 1998; Mooney, Ryan, Uhing, Reid, & Epstein, 2005; Nelson, Smith, Young, & Dodd, 1991). Three meta-analyses included experimental, quasi-experimental group design studies, and single case research (Griffith, Hurley, & Hagaman, 2009; Reddy, DeThomas, Newman, & Chun, 2009; Spencer, 2006). One meta-analysis included only quasi-experimental group studies (Ghafoori & Tracz, 2001), and two included quasi-experimental group and SCR (Joseph & Eveleigh, 2011; Ryan, Reid, & Epstein, 2004).

Of those meta-analyses that included SCR, five used PND (Joseph & Eveleigh, 2011; Mathur, Kavale, Quin, Forness, & Rutherford, 1998; Nelson, Lane, Benner, & Kim, 2011; Nelson, Smith, Young, & Dodd, 1991; and Templeton, Neel, & Blood, 2008), and four used the method described by Swanson (Swanson, & Sachse-Lee, 2000) that finds the correlation between baseline and intervention with Rosenthal's correction (Benner, Nelson, Ralston, & Mooney, 2010; Mooney, Ryan, Uhing, Reid, & Epstein, 2005; Pierce, Reid, & Epstein, 2004; and Ryan, Reid, & Epstein, 2004). One used hierarchal linear modeling (Gage, Lewis, & Stichter, 2012).

***Participant characteristics.*** Secondary studies also varied by the participants included in the primary studies. A variety of definitions for ED and EBD were used. Some also included primary studies that included smaller demographic groups.

*Disorder or disability definition.* As discussed in Chapter 1, EBD is generally a larger, more diverse population of students, members of which may have any one or a combination of emotional or behavioral problems. In contrast, ED describes a smaller group of students who qualify for special education services under the federal definition of ED. However, this discrimination is not consistently used in the secondary research reports.

Coleman and Vaughn (2000), as well as Joseph and Eveleigh (2011) used the term EBD, yet the research report indicated that only school identified students (that is, ED) were included. Many researchers included the more diverse group of students with emotional or behavioral problems (Ang & Hughes, 2001; Durlak, Furlan, & Lampman, 1991; Ghafoori & Tracz, 2001; Griffith, Hurley, & Hagaman, 2009; Maag, 2006; Mathur, Kavale, Quin, Forness, & Rutherford, 1998; McIntosh, Vaughn, & Zargoza, 1991; Nelson, Smith, Young, & Dodd, 1991; Reddy, DeThomas, Newman, & Chun, 2009; Schneider & Byrne, 1985; Scruggs, Mastropieri, & Richter, 1985; Spencer, 2006; Spencer, 2009; Sutherland, 2001). Since EBD is a broader and more vague term, one may expect these definitions to be the most varied, interfering with the ability to generalize the findings.

Still other secondary studies included students who were “at-risk” for EBD. This criterion for study inclusion broadens the participant pool further (Beelman, Pflingsten & Losel, 1994; Benner, Nelson, Ralston, & Mooney, 2010; Gage, Lewis, & Stichter, 2012; Heckaman, Conroy, Fox, & Chait, 2000; Lewis, Hudson, Richter, & Johnson, 2004; Losel & Beelman, 2003; Nelson, Lane, Benner, & Kim, 2011; and Reddy, DeThomas, Newman, & Chun, 2009).

It is possible that using the concept of EBD to get at intervention needs for students with ED will result in studies with participants whose needs are less extreme than students with ED. This is because they have not been identified as needing special education and qualifying as ED. Further, it is likely that students who are “at-risk” for ED or EBD have less deviant behaviors and emotions than students with ED. If these considerations are accurate, the inclusion of EBD and at-risk students in secondary research reports probably interferes with the ability to generalize results of that review. For example, if compared to students with ED such students have relatively minor behavior problems, they may be easier to modify, thus inflating the effect sizes.

Only a handful of secondary research efforts have reviewed research on interventions used with the more narrowly defined population of students who have been school identified with ED, as described by their state education agency (Mooney, Ryan, Uhing, Reid, & Epstein, 2005; Olmeda & Kauffman, 2003; Pierce, Reid, & Epstein, 2004; Quinn, Kavale, Mathur, Rutherford, & Forness, 1999; Ryan, Reid, & Epstein, 2004; and Templeton, Neel, & Blood, 2008). Evans, Harden, and Thomas’ (2004), and Gulchack and Lopes’s (2007) secondary studies included studies of students with behavior disorders as defined by the local culture (e.g., United Kingdom’s Department for Employment and Education).

***Demographic information.*** While most secondary researchers included all school-age students, some included narrower age or grade ranges, such as only elementary age participants (Coleman & Vaughn, 2000; Durlak, Furhman, & Lampman, 1991; Evans, Harden, & Thomas, 2004). Many of those who did include a wide age or grade range of

students aggregated data for smaller groups of students (e.g., Durlak, Furhman, & Lampman, 1991; Ryan, Reid, & Epstein, 2004; e.g., students aged 5-7 years, 8-11 years).

### **Intervention Effectiveness By Characteristic of ED**

The federal definition of ED emphasizes five qualifying characteristics, at least one of which the student must exhibit. These characteristics, stated briefly, include: (A) Inability To Learn, (B) Relationship Problems, (C) Inappropriate Behavior, (D) Unhappiness or Depression, and (E) Physical Symptoms or Fears. It is valuable to consider what interventions may be effective, and how effective, with each of these characteristics. With the understanding that there has been variation in the primary studies included in the extant secondary research, as discussed above, I now turn to the outcomes of interventions intended to affect the five characteristics of ED, as determined by secondary research.

**(A) Inability To Learn.** The first characteristic listed in the federal ED definition includes learning or “school survival” behavior problems with associated with learning (e.g., distractibility or inattentiveness; Cullinan & Sabornie, 2004). Another problem addressed under this characteristic is learning or achievement problems or delays.

***Interventions for school survival behavior.*** School survival behaviors are those that support students’ ability to learn, and include attentiveness, and following classroom routines and teacher directions (Aaroe, & Nelson, 2000). These school survival behaviors allow students to benefit from instruction. Several researchers have reviewed interventions designed to improve students’ ability to learn in school. Nelson, Smith, Young, and Dodd’s (1991) meta-analysis of self-management techniques found self-management interventions

increased on-task behavior. However, inconsistent results were found for generalization, leaving doubt about effectiveness. Sutherland (2001) examined the effects of teacher presentation rates on students with, or having behavior characteristics of students with EBD (off-task, disruptive, aggressive). Sutherland found that faster presentation rate resulted in lower rates of off-task behavior and higher participation rates. These reviews indicate interventions can improve students' school survival behaviors. Other reviewers investigate the impact of interventions on academic achievement of students.

***Interventions for academic achievement.*** Interventions aimed at improving academic outcomes for students with ED have been designed to improve academic achievement in various content areas. Ryan, Reid, and Epstein (2004) also found peer-mediated interventions, overall, to be highly effective in improving academic achievement of students with ED. More specifically, same-age peer tutoring and cross-age peer tutoring were described as highly effective. Peer-mediated interventions were found to be most effective when the student acts as both the tutor and tutee, or as the tutor only, compared to only being tutored. Peer-mediated interventions were found to be more effective when used with adolescents than with younger children (Ryan, Reid, & Epstein, 2004). Below, I review these according to the content areas of reading, writing, math, social studies and science.

***Reading.*** In their meta-analysis, Joseph and Eveleigh (2011) found that students with ED and ADHD who were taught to use self-monitoring methods performed better on reading comprehension measures than similar students who were not so taught. Griffith, Hurley and Hagaman's (2009) investigated literacy interventions for students with ED, and found

practice, direct instruction, peer tutoring, behavioral, and strategy instruction to be more effective than text modification.

Similarly, Benner, Nelson, Ralston, and Mooney (2010) found reading interventions to have an overall strong effect on reading skills of students described as at-risk for or having behavioral disorders. Joseph and Eveleigh (2011) found self-monitoring interventions (i.e., self-observation and self-recoding of behaviors) also improved the reading performance of students with EBD only when compared to those with comorbid EBD and ADHD. In addition, self-management, and peer-mediated literacy interventions were associated with larger gains. Self-instruction and multiple-component interventions were also found to have a strong effect on reading achievement (Mooney, Ryan, Uhing, Reid, & Epstein, 2005).

This was reiterated by Spencer's (2006) meta-analysis that also found peer-tutoring to be an intervention for improving reading achievement. Cross-age tutoring was found to have a large effect, reciprocal tutoring to have a medium effect, and same-age tutoring to have a small effect on reading outcomes.

Teacher-mediated antecedent academic interventions were found to be highly effective for teaching reading (Pierce, Reid, & Epstein, 2004), where as, Nelson, Lane, Benner, and Kim's (2011) "best evidence synthesis approach" found literacy instruction had small to moderate effect on literacy outcomes.

*Writing.* Self-management interventions were found to have a large effect on writing outcomes when used with students with ED (Mooney, Ryan, Uhing, Reid, & Epstein, 2005).



In contrast, cross-age peer-tutoring had a small effect on spelling when used with students with EBD (Spencer, 2006).

*Math.* The Templeton, Neel, and Blood (2008) meta-analysis identified math interventions to be effective when used with students with EBD. While EBD was not defined in their report, search terms included “behavioral disorder”, “serious emotional disturbance,” and “emotional disturbance”. Strategy instruction produced a larger effect than no strategy instruction. Surprisingly, interventions that did not include environmental accommodations (e.g., seating assignment) averaged higher than those with accommodations.

Self-management interventions were also found to have a large effect on math outcomes for students with ED (Mooney, Ryan, Uhing, Reid, & Epstein, 2005). Similarly, teacher-mediated academic interventions were found to have a large effect, with antecedent interventions more effective than consequence interventions (Pierce, Reid, & Epstein, 2004). Cross-age tutoring had less evidence of effectiveness and yielded a small effect on math outcomes (Spencer, 2006).

*Social Studies.* Three reviews targeted social studies. Self-management interventions yielded large effects on social studies achievement (Mooney, Ryan, Uhing, Reid, & Epstein, 2005). Peer-mediated interventions were shown to be effective in increasing achievement in history (Ryan, Reid, & Epstein (2004). Reciprocal tutoring was found to have a large effect on social studies outcomes when used with students with EBD (Spencer, 2006).

*Science.* Ryan, Reid, and Epstein's (2004) meta-analysis included just one study on the effectiveness of the peer-mediated interventions on science achievement. That study indicated medium gains for students with ED.

**(B) Relationship Problems.** This characteristic may be evident in students' difficulties in establishing and maintaining good relationships with peers and teachers (Cullinan & Sabornie, 2004). Social skills training is a commonly researched intervention for students with ED and EBD exhibiting relationship problems (Reddy, DeThomas, Newman, & Chun, 2009). There is extensive literature addressing reviews of social skills training and the controversies over what participant and study inclusion variables are appropriate (Cook, Gresham, Kern, Barreras, Thornton, & Crews, 2008; Quinn, Kavale, Mathur, Rutherford, & Forness, 1999).

Social skills training is not a particular intervention but a class of interventions aimed at improving social skills and social competence. Social skills training may employ various intervention techniques including coaching, modeling, rehearsal, feedback, reinforcement, goal setting, instructions, discussions, peer training, problem solving training, self-instruction, self-monitoring, self-evaluation, and self-reinforcement (Maag, 2006).

Several researchers have conducted literature reviews and meta-analyses to synthesize the extant literature on social skill instruction for students with EBD and ED. Ang and Hughes' (2001) meta-analysis of social skill training on school-aged (6- to 18 -year-old) students with antisocial behavioral problems produced a moderate effect. Similar results were found by Beelman et al's meta-analysis with students aged 3 to 15 years. Further, preschool

students (ages 3-5) demonstrated the most effect compared to lower elementary school students (ages 6-8), upper elementary school students (ages 9-11); and middle school students (ages 12-15; Durlak et al., 1991).

Others suggest student characteristics may moderate the effectiveness of social skill interventions. Ang and Huges (2001) found effects varied by the composition of the student groups participating in social skills training moderated the outcomes. Homogenous student groups had a smaller effect than those conducted with heterogeneous groupings (i.e., prosocial and deviant students). Students' characteristics have also been found to moderate the effectiveness of social skill training. The greatest effects were found for at-risk groups, followed by those with internalizing and externalizing behaviors (Beelman et al., 1994).

Further some skills showed more evidence of changed than others. High effects were found for social-cognitive skills; intermediate effects for social interaction skills; and small (but significant) effects for social adjustment measures (Beelman et al., 1994).

Some researchers were able to identify some potential sources for these varied results. McIntosh, Vaughn and Zaragoza's (1991) literature review reported that all but one study reported positive results of social skill instruction. Effects also varied by intervention theoretical approaches and models. While post intervention outcomes were relatively consistent by intervention theoretical orientation, follow-up data revealed interventions based on cognitive and cognitive behavioral theory were the stronger than those based on behavioral theory or counseling (Losel & Beelman, 2003). Operant procedures and modeling

were found to be more effective than coaching and social cognitive interventions (Schneider & Byrne, 1985).

What Works Clearinghouse, using a rigorous rubric, has analyzed the merit of branded programs for improving social outcomes. This analysis found *Coping Power* (Lochman, Wells, & Lenhart, 2004) to have potentially positive effects on social outcomes (IES, 2011a); *First Step to Success* (Beard & Sugai, 2004) to have potentially positive effects on social outcomes (IES, 2011b); and *The Incredible Years* (Webster-Stratton, 2004) to have potentially positive effects on social outcomes (IES, 2011c).

**(C) Inappropriate Behavior.** The characteristic listed in the federal ED definition as “inappropriate behavior” may include aggressive and/or disruptive acts toward peers and teachers (e.g., disrespectful or defiant behavior; Cullinan & Sabornie, 2004). Interventions for inappropriate behavior included behavioral and/or cognitive-behaviorally-orientation treatment models (Reddy, DeThomas, Newman, & Chun, 2009). Cognitive-behavioral therapy was found to reduce classroom disruptive behavior of 5- to 13-year-old students, and found to have a greater effect for those identified with conduct disorder compared to those with those a mixed diagnosis (Ghafoori & Tracz, 2001). *Coping Power* (Lochman & Well, 2002) was found to have positive effects on such behavior (IES, 2011a) when used with students with ED. Also, *First Step to Success* (Beard, & Sugai, 2004) had positive effects (IES, 2011b), and *The Incredible Years* (Webster-Stratton, 2004) was found to have potentially positive effects on such behavior (IES, 2011c) when used with students with ED.

Heckaman, Conroy, Fox, and Chait (2000) performed a systematic review of functional assessment-based interventions on students at-risk for EBD. Their studies included antecedent-based interventions, consequence-based interventions, and the combination of antecedent- and consequence-based interventions. They identified successful results for either increasing appropriate behavior or decreasing inappropriate behavior.

**(D) Unhappiness or Depression and (E) Physical Symptoms or Fears.** (D) Unhappiness or Depression describes students' unpleasant thoughts and feelings, as evidenced in verbal and other behavior (e.g., expresses little pleasure or joy). (E) Physical Symptoms or Fears relates to unpleasant thoughts and feelings involving anxiety and physical distress (e.g., worried, tense). All mental, emotional, and behavior problems have a private component to some extent, but these two kinds of problem are largely private. This not only presents challenges of assessment, it means that students with such problems do not bother adults and peers as much as the readily visible disorders such as poor achievement and disruptive conduct. At present, no secondary research could be located that addressed either of these two characteristics.

### **Conclusion**

Secondary research, especially that involving systematic reviews and meta-analyses, is a critically important endeavor in understanding interventions for students with ED. However, there are significant obstacles to using systematic reviews and meta-analyses in ways that will obtain optimal information from the primary research on which they are based. Researchers have used a variety of primary research methods including, group comparison

and SCR designs, to evaluate the effectiveness of such interventions. In addition, researchers have also a variety of analysis methods to synthesize the primary research.

Primary research studies that use SCR dominate research on interventions for students with ED. Meta-analysis of this research is complicated by uncertainty as to what should constitute ES. Also, although many researchers have investigated aspects of interventions for students with ED, none have investigated interventions effects according to the five characteristics of ED in the federal definition. In fact, two of these characteristics of ED appear to have been missed by systematic reviews and meta-analyses. Yet the most successful interventions for students with ED may be those that match the presenting problem (Peacock Hill Working Group, 19991). The present dissertation study addresses this and other issues in secondary research on interventions for students with ED. It is a meta-analysis designed to identify which interventions are effective for students with ED, and how various interventions' effectiveness is moderated by the characteristics of ED.

### CHAPTER 3

Chapter 3 describes the methodology for addressing the research questions set out in Chapter 2 (a) Do interventions improve outcomes for students with ED? (b) Are ED characteristics related to the effectiveness of interventions? (c) Which interventions are most effective for students with different ED characteristics? To pursue these questions I first identified appropriate published single-case research (SCR) intervention studies that include students with emotional disturbance (ED). This involved a search strategy followed by an exclusion strategy. Second, I coded the surviving studies according to a variety of study, participant, intervention, dependent variable, and other features.

Third, I compared phases of the SCR studies that were relevant to answering questions about the value of interventions to produce changes in the dependent variables under study. To make these comparisons I used an index of data point nonoverlap called *PAND plus robust Phi*. Both Percent of Non-Overlapping data (PAND) and Robust Phi (RPhi) quantify how confident one should be as to whether the two sets of data in two compared phases are actually different from one another. Each phase comparison was assigned a PAND value, calculated in a standard way. From the PAND value, RPhi and related values were derived.

Fourth, I examined these PAND values and their derivatives via descriptive and inferential statistics. Initially, I aggregated the PANDs across all phase comparisons to yield an omnibus “overall PAND”. Next, I disaggregated the PANDs and their derivatives

according to various interesting contrasts (e.g., different SCR design, different characteristics of ED).

### **Search Method**

Studies were identified for inclusion in this meta-analysis by means of a thorough search of computer databases and a hand-search of selected journals. Subsequently, the reference lists of identified studies were reviewed in search of studies that eluded previous search attempts.

First, an electronic search using the key words “emotion\* dis\* and behavior\* dis\*” using ERIC (Ebscohost), PsychInfo (Ebscohost) and SocAbstracts (Ebscohost) was conducted. The asterisks allowed for a search that included the variations of those terms (i.e., emotional disturbance, emotionally disabled, emotionally disturbed, behavior disorders, etc.). Next, a hand search was performed in the following journals deemed likely to contain relevant studies: Behavioral Disorders, Education and Treatment of Children, Exceptional Children, Journal of Abnormal Child Psychology, Journal of Applied Behavior Analysis, Journal of Behavioral Education, Journal of Child Psychology and Psychiatry, Journal of Emotional and Behavioral Disorders, Journal of School Psychology, Journal of Special Education, Psychology in the Schools, School Psychology Quarterly, and School Psychology Review. Finally, ancestral searches (i.e., review of reference lists) of identified articles and relevant reviews were undertaken.



### **Study Inclusion and Exclusion**

An eligible study was an original single-subject experiment that included the manipulation of at least one independent variable (i.e., treatment) and measured at least one outcome variable. The study must have occurred in the United States or Canada, and have been published in English in a peer-reviewed journal between 1984 and 2010 inclusive. The study must have been conducted in a school or other educational environment; the participants must have been students aged 5 to 21 inclusive, identified as having either ED or a “qualifying EBD.”

**Eligible treatments.** Eligible treatments included psychological, educational, or behavioral interventions involving actions performed with the expectation that they would have beneficial effects on the students receiving them. Eligible interventions must have been performed in an educational setting on eligible students.

**Outcome variable eligibility.** An eligible outcome variable was one that (a) was expressed in terms of quantitative data about students’ behavioral, cognitive, academic, social, emotional, personal, or other eligible functioning; and (b) measured effects of an eligible treatment (but not drug or surgery); and (c) included data from which an effect size could be derived. The following were ineligible outcome variables: (d) student variables that were impossible or unlikely to be changed by the treatment investigated, such as personal and demographic characteristics (e.g., age, sex, race-ethnic status); (e) medical or physical variables (e.g., weight, respiration, levels of hormones) unless the study clearly tied such variables to student behavioral or emotional functioning; (f) functioning of persons other than

the student, such as teachers, parents, siblings, peers, etc.; or (g) outcome variables measured out of school.

**Experimental research design.** To be included in this meta-analysis, the SCR design must have been: (a) withdrawal (or reversal), (b) alternating treatments (c) simultaneous treatments, or (d) multiple baseline designs (across participants, settings, behaviors, or probes [very brief tests of specific academic material]). Also, the SCR design must have included the manipulation of at least one independent variable. Each SCR design must have included at least one baseline phase.

Simple AB (i.e., one baseline and one treatment phase) designs were not included. SCR designs that were not experimental (i.e., did not manipulate an independent variable and measure a dependent variable) were not included.

**Location, and publication.** As noted before, studies must have been conducted in the United States or Canada, in a school. Studies had to have been published in a peer-reviewed journal, in English between the years of 1984 and 2010 inclusive.

**Participants.** To be included in this meta-analysis, the SCR must have included one or more participants aged 5 to 21 inclusive, or in kindergarten through twelfth grade. Study participants must have been students attending a public or private school, or other educational institution.

***ED or eligible EBD.*** In addition, the participant(s) must have been eligible under IDEA in the category of ED (or equivalent state term) and receiving special education (or with an active IEP) in any education environment. Studies were still eligible if (a) the

participants were identified as ED in addition to learning disabilities, other health impairment, or attention deficit hyperactivity disorder; or (b) participants had been officially classified as having a mental disorder found in the DSM of an emotional or behavioral nature, and were receiving education in any education environment; or (c) participants were not classified as ED or with a DSM disorder, but were described as having emotional or behavioral problems, and receiving education in a restrictive education environment (e.g., separate classroom in regular school, separate or alternative school, residential placement) for students with behavior or emotional problems.

Studies were not eligible if participants were students not identified as ED even if identified with another education disability. Consistent with the federal definition of ED, Socially Maladjusted students were included only if they were also identified as ED.

**Exclusions.** Studies were excluded if they did not meet the above criteria for inclusion, or if data for eligible participants could not be separated from data for ineligible ones because such individual data were combined to yield group data (e.g., for a classroom).

### **Coding Procedure**

Operational definitions of the variables to be coded and a coding form were developed to record information contained in the SCR research reports. The author coded each of the studies, using a coding procedure expanding on that reported by Skiba and Casey (1985). The following criteria were used for coding purposes (See Table 1).

**Participants and setting.** The following items related to study participants were coded:

***ED label.*** The label used to describe the emotional disturbance of the participants was coded (e.g., serious emotional disturbance, behavioral disorder). Additional disabilities, if any, were coded as well (i.e., LD or OHI in addition to ED).

***ED characteristics and features.*** The five characteristics found in the federal definition of ED were coded (see page 5); briefly stated these are: (A) Inability to Learn, (B) Relationship Problems, (C) Inappropriate Behavior, (D) Unhappiness or Depression, and (E) Physical Symptoms or Fears. In addition to characteristics, Socially Maladjusted and Schizophrenia-Like Features were coded as well, since these terms are also found in the federal definition of ED. Each characteristic and feature was defined based on the behaviors and emotions presented in the Teacher Rating Form and the Observation Form of the Scales for Assessing Emotional Disturbance-2, for which there is evidence of content, criterion-related, and construct validity (Epstein & Cullinan, 2010). These five characteristics and two features of ED were used in moderator sub-analyses. They are described in more detail as follows.

*(A) Inability to Learn.* If a study described a qualifying student's behaviors as any of the following, the student was coded as (A) Inability to Learn: poor academic performance (e.g., math, reading, spelling); weak listening or note-taking skills; inadequate ability to complete independent work; easily distracted (e.g., does not pay attention to teacher or work); lacks interest, motivation, or positive attitude toward school tasks; or dependency on adults or peers to complete school tasks. Qualifying students diagnosed with attention deficit

hyperactivity disorder (or equivalent term) were also coded as having the Inability to Learn characteristic.

*(B) Relationship Problems.* If a study described a qualifying student's behaviors as any of the following, the student was coded as (B) Relationship Problems: has few friends; does not work well in a group; rejected or avoided by peers; picked on or persecuted by other students; lacks social skills; or avoids interacting with people.

*(C) Inappropriate Behavior.* If a study described a student's behaviors as any of the following, the student was coded as (C) Inappropriate Behavior: disrespectful or defiant of authority; cruel to peers; verbally aggressive, teases or taunts others; fails to consider the consequences of own acts; cheats, lies or steals; makes threats; is loud, disruptive or rowdy; destroys or ruins things; physically assaults or fights; or uses obscene, profane, or sexually oriented language. Qualifying students diagnosed with oppositional defiant disorder or conduct disorder were also coded as exhibiting Inappropriate Behavior.

*(D) Unhappiness or Depression.* If a study described a qualifying student's behaviors as any of the following, the student was coded as (D) Unhappiness or Depression: lacks self-confidence; has feelings of worthlessness; experiences little joy or pleasure; sad much of the time; rarely smiles; little or no interaction with teachers; loss of interest in formerly enjoyed experiences; pessimistic about the future; makes statements that indicate sadness, dejection, despair or gloom; or appears to lack self-confidence.

*(E) Physical Symptoms or Fears.* If a study described a qualifying student's behaviors as any of the following, the student was coded as (E) "Physical Symptoms or Fears":

complains of physical discomfort; anxious, worried, or tense; afraid of unlikely dangers or calamities; talks of suicide or own death; overly sensitive feelings and emotions; feels excessively guilty; shows nervous habits (e.g., tics, bites nails, twists hair); or harms own body (e.g., picks self, cuts self, writes on self).

*Socially Maladjusted.* If a study labeled a qualifying student as being socially maladjusted, or described the students' behaviors as any of the following, the student was coded as Socially Maladjusted: vandalizes property outside of school; participants in illegal, anti-social, or gang activities outside of school; abuses drugs or alcohol before or after school; exhibits precocious sexual behavior outside of school; runs away from home; or steals outside of school.

*Schizophrenia-Like Features.* If a study described a qualifying student's behaviors as any of the following, the student was coded as Schizophrenia-Like Features: schizophrenic or having schizophrenia, thought problems, or psychotic.

*Age and grade.* Study participants' age and school level were also coded. If only age was provided, those in Kindergarten were assumed to be 5 years old, first grade 6 years old, and so forth. If only age was reported, the same conversion was conducted in reverse.

*Sex.* Study participants' sex was noted. Sex was coded as either male, female, or not determinable.

*Race.* Study participants' race was collected. Race was coded as African American, Caucasian, Hispanic, other, or not determinable.

**Placement.** The educational placement of the study participants was coded as public school, private school, or unreported. Special education placement was also coded in a manner consistent with the continuum of services described by IDEA; (a) Regular is defined as receiving services less than 21% out of regular class; (b) Resource, 21%—60% out of regular class; (c) Self-Contained, greater than 60% out of regular class but in regular school; (d) Alternative, greater than 50% of day in alternative or day-treatment school; (e) Residential or hospital setting; or (f) Unreported. Settings described as “separate program” were coded as Self-Contained, unless they were described as being located in a separate building or facility, in which case they were coded as Alternative.

**Study characteristics.** The following items related to the characteristics of the studies were coded:

**Study design.** The SCR design was coded as: (a) withdrawal (or reversal), (b) alternating treatments (c) simultaneous treatment, or (d) multiple baseline (MB) designs. Further, the type of multiple baseline design was also coded as: (d<sub>1</sub>) across settings, (d<sub>2</sub>) across participants, (d<sub>3</sub>) across probes (short academic tests), or (d<sub>4</sub>) across behaviors.

**Interobserver agreement and quality of evidence.** What Works Clearinghouse (WWC: Kratochwill, Hitchcock, Horner, Levin, Odom, Rindskopf, et al., 2010) describes SCR design evidence standards as follows:

- (a) Each outcome variable must be measured systematically over time by more than one assessor, and the study needs to collect inter-assessor agreement in each phase

and on at least twenty percent of the data points in each condition (e.g., baseline, intervention) and the inter-assessor agreement must meet minimal thresholds.

(b) The study must include at least three attempts to demonstrate an intervention effect at three different points in time or with three different phase repetitions.

(c) For a phase to qualify as an attempt to demonstrate an effect, the phase must have a minimum of three data points (meets evidence standards with reservations) or five data points (meets evidence standards). (WWC, 2010, p.15.)

These WWC IOA criteria suggested the need to code IOA in three ways. The first recorded whether or not the study reported IOA checks on 20% or more of the observations. The second recorded whether or not the study reported IOA check of 20% or more of the observations in each phase. The third recorded whether or not the IOA agreement met the minimum criteria for accuracy (see Kratochwill, et al, 2010, or Suen & Ary, 1989 for thresholds).

In addition, studies were coded as to the degree in which it met the WWC evidence standards, using their criteria. Studies must have reported systematically manipulating an independent variable, and systematically recording outcome variables over time. Further, studies must have sufficient phase repetitions, as well as a sufficient number of data points per phase. These criteria differ by SCR design and are as follows.

A withdrawal design was considered to have “met evidence standards” if it contained at least four different phases (e.g., ABAB), with at least five data points per phase. If the



withdrawal design had at least four phases with at least three data points per phase, it was coded as “met evidence standards with reservations.”

A multiple baseline design was considered to have met evidence standards if it included at least six different phases (e.g., AB for three different participants), with at least five data points per phase. If the multiple baseline design had at least six phases with at least three data points per phase, it was coded as met evidence standard with reservations.

Alternating and simultaneous treatment designs were coded as met evidence standard if each treatment included total of at least five data points, and coded as met evidence standard with reservations if each treatment included at least three data points.

***Outcome variables.*** Outcome variable categories were coded as either behavioral or academic. Behavioral outcomes were those that used target-behavior recording, teacher ratings, or other methods to measure social or personal behaviors (e.g., hitting, negative statements, hand raising). Academic outcomes included achievement probes and similar academic outcomes, and learning behavior outcomes (e.g., on-task behavior). When more than three dependent variables were measured, only the first three outcomes presented were coded.

***Interventions.*** Interventions were coded as either behavior/emotional or academic depending on which outcome was intended to change. These interventions codes are not mutually exclusive. Some interventions may be used to change behavioral as well as academic outcomes. In some cases SCR included multiple interventions. When this was the case, only the first two intervention types were coded.

***Behavior/emotional interventions.*** Behavior/emotional interventions are those that aim to change students' behavior or emotional response. The following kinds of behavior/emotional interventions were coded: (a) choice (e.g., students had a choice of reinforcement); (b) use of classroom procedures (i.e., routines); (c) function-based interventions (i.e., interventions developed to meet the need or function of the undesirable behavior); (d) positive reinforcement (i.e., the delivery of a pleasant stimulus [e.g., praise, desirable activity, and edibles] contingent on a desired behavior); (e) negative reinforcement (i.e., the removal of an unpleasant stimulus contingent on a desired behavior); (f) punishment (i.e., the application of an unpleasant stimulus, or removal of a pleasant stimulus, contingent on an undesired behavior); (g) self-management/monitoring (e.g., self-reinforcement, self instruction, self-talk, mindfulness); (h) peer modeling (i.e., live peer demonstrating a desirable behavior); (i) video modeling (i.e., viewing the recorded demonstration of a desirable behavior); and (j) social skill instruction (i.e., instruction aimed at improving social skills). Token economies with response cost systems were coded with two codes: positive reinforcement and punishment.

***Academic interventions.*** Academic interventions are those that aim to change students' academic behavior (e.g., math scores, time on-task). Academic interventions were then coded in the following categories: (a) instruction (teaching or education); (b) computer-assisted instruction (teaching or education with a computer); (c) direct instruction (explicit, structured teaching of a skill with immediate corrective feedback and continuous progress monitoring); (d) opportunities to respond (the provision of multiple occasions for student

response); (e) curricular modifications (adjustment of educational activities/materials); (f) learning strategy instruction (explicit teaching of learning processes); and (g) peer tutoring (instruction by a student).

**Inter-observer agreement coding.** Inter-observer agreement for coding articles was assessed by the researcher and her committee chair, a committee member, and two other students enrolled in the doctoral program. They coded and recorded data independently on approximately 20% of randomly selected eligible studies. Inter-observer agreement was calculated by dividing the number of agreements by the sum of the number of agreements and disagreements, and then multiplying the quotient by 100. The following participant variables were coded: (a) ED label, (b) ED characteristic, (c) participants age and/or grade, (d) sex, (e) race, (f) educational placement. The following study variables were coded: (a) single case research design, (b) quality of evidence, (c) outcome variables, and (d) intervention type. Finally, IOA was also conducted on the PAND calculations, including: (a) number of data points in baseline and (b) number of data points in treatment phases, and (c) percent of non-overlap. See Table 1 for the codification of independent and dependent variables.

### **Data Analysis**

Three levels of data analysis were employed in this meta-analysis. The first analysis was the individual phase comparison-level, followed by study-level analysis. For these, PAND, RPhi and 95% confidence intervals were calculated. The final level analysis disaggregated groups of phase comparisons according to interesting moderator variables

(e.g., kind of intervention or SCR design). Overall PAND, Overall RPhi and confidence intervals were calculated for the third level of analysis.

**Statistical independence.** Consistent with commonly used meta-analytic literature review guidelines to reduce statistical dependencies (Cooper & Hedges, 1996), samples reported across more than one study would have been excluded to ensure the independence of the samples. No duplicated samples were found; therefore all included studies represented independent samples with unique participants.

**Phase comparison rules.** Individual phase contrasts (level 1) were conducted as described in Parker and Hagan-Burke (2007), and Schneider, Goldstein, and Parker (2008) according to these rules. (a) When the same intervention is repeated in multiple treatment phases, all intervention phases were combined and all baseline phases were combined. For example in  $A_1B_1A_2B_2$  designs,  $A_1$  was added to  $A_2$  and compared to  $B_1$  plus  $B_2$ . In  $A_1B_1A_2B_2A_3B_3$  designs,  $A_1$  was added to  $A_2$  and  $A_3$  and compared to  $B_1$  plus  $B_2$  plus  $B_3$ .

(b) When differing interventions were employed in treatment phases, each intervention phase data were compared to the preceding baseline data. In  $A_1B_1A_2C_1$  designs,  $B_1$  was compared to  $A_1$  and  $C_1$  was compared to  $A_2$ . This type of design yields two comparisons; the first compares the first treatment to the preceding baseline, and the second compares the second treatment to its preceding baseline.

(c) For alternating and simultaneous treatment designs, all data points of the same intervention were combined and compared to the baseline (i.e., no-treatment condition). Designs of this type yield one comparison for each treatment. For example, an alternating

treatment design that alternates among three different interventions (and includes the required baseline condition) yields three comparisons.

(d) In designs that employed a variation of withdrawal and alternating treatment designs (e.g.,  $A_1B_1C_1A_2B_2A_3C_2$ ), treatment phases were coded and compared to the previous baseline phase consistent with (b) described above. In other words,  $A_1$  was grouped with  $A_2$  and  $A_3$  and compared to  $B_1$  grouped with  $B_2$ . The  $C_1$  and  $C_2$  phases were combined and compared to the combination of all three baseline phases.

(e) For multiple baseline designs the rules above were applied separately to each row. A multiple baseline across participants, with three participants and one treatment, would yield three comparisons (e.g., B to A, for all three participants). The same design with two treatments beyond the baseline would yield six comparisons (e.g., B to A and C to A, for all three participants).

While these individual phase comparisons and the related PAND and RPhi calculations allow comparison of phases, the forthcoming Overall PAND, Overall RPhi method allows computation of CI for for groups of comparisons. The availability of CIs permits the contrasting of effects according to independent (moderator) variables.

### **Estimation of Effect Sizes**

Three nonoverlap-related statistics were used to examine the research questions. These are Overall PAND, Overall RPhi, and confidence interval (CI) for Overall RPhi. Overall PAND is an extension of PAND (Schneider, Goldstein, & Parker, 2008; J. Davis,

personal communication, 1/16/2012). While PAND quantifies nonoverlap for one SCR study, Overall PAND summarizes nonoverlap for any group of SCR studies.

**PAND calculation.** Percent of all non-overlapping data for a single study is calculated as the percentage of data remaining after removing from a SCR study's graph the fewest data points that would eliminate all overlap of data points across phases (Parker, Hagan-Burke, & Vannest, 2007). The formula for PAND is as follows:

$$\text{PAND} = \frac{\text{Number of data points after removal of overlap}}{\text{Total number of data points in all phases}}$$

For example, suppose a SCR study has 10 data points total in two baseline phases, and 15 data points in two treatment phases, with 4 data points overlapping. This produces a PAND of  $(10+15-4) / (10+15) = 21/25 = .84$ . Expressed as a percentage, this is 84% nonoverlap.

**RPhi calculation.** The RPhi is calculated from PAND calculations, using a  $2 \times 2$  table as follows:

A	B
C	D

In this table,  $A = D$  = the number of data points removed to eliminate all overlap, divided by 2.  $B$  is the number of data points left in the treatment phase(s), and  $C$  is the number of data points left in baseline phase(s). The formula for RPhi is as follows:

$$R\text{Phi} = \frac{(BC) - (AD)}{\sqrt{(A+B)(C+D)(A+C)(B+D)}}$$

Applied to the sample data set for calculating PAND, described above, the result is:

$$R\text{Phi} = \frac{(13*8) - (2*2)}{\sqrt{(2+13)(8+2)(2+8)(13+2)}} = 0.667$$

**Confidence intervals.** Confidence intervals and p-values for RPhi were calculated with a web-based calculator ([javapages.org](http://javapages.org)) that uses a general method for 2x2 contingency tables. The method involved the following, from J. Pezzullo, creator of the calculator:

finding two limiting 2x2 tables that represent the confidence limits of the tables themselves. The two limiting tables are obtained using an iterative root-finding algorithm, replacing expected values with the upper or lower limiting table in the chi-square contingency-table computation, to get a chi-square value of 3.88 (the critical chi-square value for  $p=.05$ ; Pezzullo, personal communication, 10/3/2011).

The statistical significance of a single RPhi value can be judged as follows: if its confidence interval does not include zero, it is significantly different from zero. A pair of RPhi values can be compared using their confidence intervals as well: they are significantly different if their respective confidence intervals do not overlap (May, 2003).

**Overall PAND calculation.** Overall PAND calculation is an extension of PAND calculation, but applied to a group of SCR comparisons instead of a single SCR comparison.

For the group of SCR comparisons, number of baseline phase data points for each SCR comparison are summed, number of treatment phase data points for each SCR comparison are summed, and number of data points removed to eliminate overlap are summed. Then the PAND formula (above) is applied to these sums to yield Overall PAND (Schneider, Goldstein, & Parker, 2008; J. Davis, personal communication, 1/16/2012):

$$\frac{(\text{sum of baseline data points} + \text{sum of treatment data points}) - \text{sum of removed data points}}{(\text{sum of baseline data points} + \text{sum of treatment data points})}$$

For example, suppose a group of SCR comparisons has a total sum of 1000 data points in all baseline phases, and 1500 data points in all treatment phases, with a sum of 300 data points removed because they eliminated overlap. This produces an Overall PAND of  $(1000 + 1500 - 300) / (1000 + 1500) = 2200/2500 = .88$ . Expressed as a percentage, this is 88% nonoverlap for the group of SCR study comparisons.

**Overall RPhi calculation.** Analogously to RPhi, Overall RPhi is calculated from Overall PAND calculations, using a  $2 \times 2$  table as follows:

A	B
C	D

In this table, A = D = the sum of all data points removed to eliminate overlap (summed across all like phases in the aggregate group), divided by 2. B is the sum of data points left in the treatment phases (sum of all treatment phase data points in the aggregate group minus half of the overlap), and C is the sum of data points left in baseline phases (sum



of all baseline phase data points in the aggregate group minus half of the overlap). The formula is as follows:

$$\text{Overall RPhi} = \frac{(BC)-(AD)}{\sqrt{(A+B)(C+D)(A+C)(B+D)}}$$

**Confidence intervals and significance of Overall RPhi.** Confidence intervals (95%) for Overall RPhi were calculated as described above, with the web-based calculator (javapages.org). The statistical significance of an Overall RPhi value was judged as follows: if its confidence interval does not include zero, it is significantly different from zero. Pairs of Overall RPhi values were compared using their confidence intervals as well: a pair is significantly different if their respective confidence intervals do not overlap.

**Aggregating and disaggregating effect sizes.** All data from the included studies were aggregated. Then subsets of data were disaggregated by selected coding variables described above anticipated to be potential moderators. Overall PAND and RPhi were then calculated for all included studies as well as for moderating variables.

**Overall PAND calculation.** Overall PAND was calculated for each analysis by dividing the sum of all non-overlapping data points, by the total number of data points in all comparisons included in the analysis (Schneider, Goldstein, & Parker, 2008).

**Overall RPhi calculation.** Overall RPhi was calculated with the RPhi formula (described above) with the results of the Overall PAND procedure for all comparisons included in the analysis. The related confidence intervals were also calculated on these data.

**RPhi confidence intervals.** Confidence intervals provide a measure of the precision of the effect size and were calculated using the aforementioned web-based calculator

(javapages.org). For this meta-analysis, conservative 95% confidence intervals were calculated, which align with the p-value of 0.05%, and allows for significance testing ( $\alpha = 0.05$ ).

**Interpretation guidelines.** Interpretation guidelines vary by the type of effect size, and generally include categories of “small”, “moderate”, or “large”. Phi coefficients less than 0.20 are considered negligible, 0.20-.49 small, 0.50-.79 moderate, and 0.80 or greater as strong (Cohen, 1988).

## CHAPTER 4

In this chapter, results are organized by the three research questions of this study, which were to determine whether and to what extent: (a) interventions improve outcomes for students with ED, (2) ED characteristics moderate the effectiveness of interventions, and (3) which interventions are most effective for students with different ED characteristics.

In order to present the results of this meta-analysis responding to the research questions listed above, I describe the primary studies, and participants included. I present descriptive statistics on study characteristics including publication source and year, research design, interventions employed, and outcome measures. Then I present intervention effects, via PAND, and RPhi with corresponding confidence intervals for each of the independent variables analyzed. PAND and (RPhi) quantify the extent two sets of data in two compared phases are actually different from one another. Confidence intervals indicate the precision of effect-size measurement. Finally, intervention effects are analyzed by ED characteristic (e.g., Inability to Learn, Relationship Problems). For each characteristic, each type of intervention is analyzed.

### **Studies Included**

The search described in Chapter 3 led to the identification of 186 studies of which 73 studies met the inclusion criteria. See Table 2 for a summary of all included studies. Number of studies and number of comparisons are denoted,  $k$  = number of studies, and  $n$  = number of comparisons). See Table 3 for PAND and RPhi analysis for each study included, as well for every comparison.

**Studies excluded.** Of the 186 studies identified in the search, 113 studies were excluded for the following reasons (see Appendix C for the list of all excluded studies):

1. The study did not investigate an intervention ( $k = 29$ ).
2. The study participants did not meet the definition of ED ( $k = 28$ ).
3. The study did not occur in a school environment ( $k = 27$ ).
4. The study did not investigate student outcomes ( $k = 20$ ).
5. The study did not take place in the United States or Canada ( $k = 9$ ).

The remaining 73 research reports met the criteria and were included in this meta-analysis. Because two reports included two experiments each, a total of 75 experiments were included. Because many experiments contained multiple comparisons and multiple outcome measures, the number of comparisons for this meta-analysis totaled 713.

**Included studies' features.** Of note are some details about the studies included in this meta-analysis, especially publication year, journal, and research design.

**Publication years.** The studies included were most likely to be published in the years of 1995 ( $k = 7$ ) and 2005 ( $k = 7$ ). Overall the number of SCR studies published is increasing over time. Table 4 contains the number of studies published each year.

**Journals.** Forty percent of all studies were published in *Behavioral Disorders* ( $k = 11$ ), *Journal of Emotional & Behavioral Disorders* ( $k = 9$ ) and *School Psychology Review* ( $k = 9$ ). The distribution of journals containing included SCR studies can be found in Table 5.

**Research designs.** More of the SCR studies included in this meta-analysis employed the multiple baseline across subjects design ( $k = 29$ ), ABAB ( $k = 11$ ), alternating treatments

( $k = 7$ ), multiple baseline across probes ( $k = 7$ ), and multiple baseline across settings ( $k=7$ ). Other designs included: complex variations of the ABC design ( $k = 8$ ), and multiple baseline across behaviors ( $k = 5$ ).

**Participant characteristics.** Seven hundred nine participants were included in all the studies. The following describes participant included in this analysis.

**Sex.** Of the participants for whom sex was reported, 554 (86%) were male and 87 were reported as female. The sex of 68 participants was not reported.

**Race.** The racial background was not reported for most of the participants ( $n = 429$ ). Of those for which race was reported, most were Caucasian ( $n = 135$ ) and African American ( $n = 126$ ). Twenty-three were reported as Hispanic.

### **Inter-observer Agreement on Coding**

On 20% of the included studies, inter-observer agreement (IOA) was calculated for each variable coded, using the formula ( $[\text{number of agreements} / \text{the sum of the number of agreements plus disagreements}] * 100$ ). Overall the IOA ranged from 78% to 100%. The range for IOA for study related variables was 88 to 100% (research design = 96%, journal title = 100%, year published = 100%, unit of analysis = 100%, setting = 88%). The range for IOA for study participant demographic data was 78 to 100% (participant sex = 100%, race = 100%, age = 100%, grade 100%, ED term = 100%, ED characteristic A = 78%, ED characteristic B = 100%, ED characteristic C = 99%, ED characteristic D = 92%, ED characteristic E = 95%, Schizophrenia-Like Features = 96%, and Socially Maladjusted = 100%). The range for IOA for research report variables was 91 to 100% (unit of analysis =

100%, treatment fidelity reported = 100%, study quality (WWC criteria) = 97%, number of data points in baseline phases = 100%, number of data points in treatment phases = 100%, and percent of non-overlap = 91%). All IOA results were within acceptable ranges.

### **Effect sizes**

**Overall PAND and Overall RPhi.** The Overall PAND for all comparisons included in this meta-analysis was 87.2%, and the Overall RPhi coefficient was 0.744 (95% CI = 0.732, 0.755). These data indicate that interventions used with students with ED result in a large effect.

**Outcomes by participant sex and race.** As Table 6 represents, there was no statistical difference between outcomes for males and females (when sex was reported). However, participants for whom sex was not reported demonstrated significantly larger outcomes. Of the participants for whom race was reported, Caucasian and Hispanic participants demonstrated significantly larger effects than African American participants and those whose race was not reported.

**Outcomes by ED characteristics.** Table 7 summarizes the disaggregated findings regarding the five characteristics of ED and *Schizophrenia-Like Features* and *Socially Maladjusted*. Recall the ED definition and the five characteristics from Chapters 1 and 2. Also notice that the federal definition of ED notes schizophrenia and social maladjustment. As such, these are also included in the analysis. In the column headed “Confidence Interval” entries for which the confidence limits do not overlap are significantly different from each other.

*(A) Inability to Learn.* Students coded as exhibiting Inability to Learn (A) had a significantly larger Overall RPhi than students with ED characteristic (B), (D) and (E). Conversely those with characteristic (A) had a significantly lower Overall RPhi than students with Schizophrenia-Like Features.

*(B) Relationship Problems.* Students coded as exhibiting ED characteristic (B) had a significantly lower Overall RPhi than students with ED characteristic (C), as well as those with Schizophrenia-Like Features and Socially Maladjusted students .

*(C) Inappropriate Behavior.* Students coded as exhibiting ED characteristic (C) had significantly higher Overall RPhi than students with ED characteristics (D) and (E).

*(D) Unhappiness or Depression.* Students coded as exhibiting ED characteristic (D) had a significantly lower Overall RPhi than students with Schizophrenia-Like Features.

*(E) Physical Symptoms or Fears.* Students coded as exhibiting ED characteristic (E) had a significantly lower Overall RPhi than those with Schizophrenia-Like Features and Socially Maladjusted students.

*Schizophrenia-Like Features.* Students coded as exhibiting Schizophrenia-Like Features had a significantly higher Overall RPhi than students with ED characteristics (B), (D), (E) and Socially Maladjusted students.

*Socially Maladjusted.* Students coded as Socially Maladjusted had a significantly higher Overall RPhi than students with characteristics (A), (B), (C), (D), and (E). Socially Maladjusted students had significantly lower Overall RPhi than those with Schizophrenia-Like Features.

**Setting.** Table 8 presents comparisons by setting. More than half the comparisons occurred in Self-Contained settings ( $n = 381$ ). Alternative schools, residential schools, day treatment facilities, and psychiatric facilities, were collapsed into one category called “All Separate”, and accounted for 144 of the comparisons. More than twice as many comparisons occurred in the resource setting ( $n = 78$ ) than in the general education setting ( $n = 33$ ).

Again, in the column headed “Confidence Intervals”, pairs of Overall RPhi entries for which the confidence intervals do not overlap are significantly different from each other ( $p < .05$ ). General Education setting was not significantly different from Resource, but both were significantly higher than Self-Contained and All Separate. Moreover, Self-Contained was significantly lower than All Separate.

**Outcomes by study characteristics.** Tables 8 and 9 present the Overall PAND, Overall RPhi coefficient, and 95% confidence intervals for outcomes by SCR design, and research fidelity and quality.

**Analysis by SCR design.** As can be seen in the Table 9, multiple baseline across probes was associated with the largest effect sizes, and complex ABC designs produced the smallest effect sizes.

**Analysis by research fidelity and quality.** Table 10 presents the overall PAND, Overall RPhi coefficient, and 95% confidence intervals of studies that did report implementation fidelity and quality indicators. More of the comparisons included in this meta-analysis originated from research reports that did not describe intervention fidelity measures ( $n = 431$ ), than did ( $n = 281$ ). Those reports that did report intervention



implementation fidelity measures produced larger effects (PAND = 89.8%, RPhi = 0.794 [95% CI = 0.776, 0.811]) than those that did not (PAND = 85.7%, RPhi = 0.715 [95% CI = 0.699, 0.729]).

***Analysis by What Works Clearinghouse criteria.*** Of those SCR comparisons included in this analysis, fewer than one third of them met the WWC IOA criterion ( $n = 211$ ). Of those, six did not meet the evidence standards; 107 met evidence standards for phase repetitions and number of data points per phase with reservations; and 98 met evidence standards. While the sample size of comparisons that did not meet evidence standards was small, that sample produced significantly smaller effects. There were no significant differences in the ES for those comparisons that originated in studies that met the standards with reservations versus those that met the standards without reservations.

Seventy-five comparisons originated from studies that either did not report IOA or reported an IOA based on less than 20% of all observations. Most of the comparisons originated from reports that reported IOA procedures using terms that were less specific than those required by the What Works Clearinghouse. For example, in a report indicating that “IOA was conducted on 23% of observations, with at least one observation occurring in each phase” one cannot be certain that IOA checks were conducted on at least 20% of observations in each phase. Vague wording resulted in some studies being coded as not meeting the What Works Clearinghouse criteria.

***Analysis by intervention type.*** As discussed earlier, some primary studies investigated one type of intervention, while others employed combinations of interventions.

Table 11 presents the Overall PAND, Overall RPhi, and 95% confidence intervals for intervention only (i.e., “intervention implemented in isolation”), interventions used in combination (e.g., positive reinforcement plus another intervention), as well as the total (i.e., intervention only plus intervention in combination) for each intervention type.

***Self-Management/Monitoring.*** Twenty-two studies, containing 225 comparisons, evaluated Self-Management/Monitoring. More of those comparisons investigated the use of this intervention in isolation ( $n = 174$ ) than in combination with another intervention type ( $n = 51$ ). Self-Management/Monitoring used in isolation resulted in a significantly larger effect than when it was used in combination. In total, Self-Management/Monitoring yielded larger effects ( $p < .05$ ) than Curricular Modifications and Choice, and smaller effects than Instruction (See Table 11).

***Positive Reinforcement.*** Sixteen studies, containing 193 comparisons, evaluated Positive Reinforcement. More of those comparisons investigated the use of this intervention in combination ( $n = 123$ ) than in isolation ( $n = 70$ ). Positive Reinforcement used in isolation resulted in a significantly larger effect than when it was used in combination. In total, Positive Reinforcement yielded larger effects than Curricular Modifications, Choice, and smaller effects than Instruction.

***Instruction.*** Eleven studies evaluated Instruction. They contained 119 comparisons. More of those comparisons investigated the use of this intervention in isolation ( $n = 90$ ) than in combination ( $n = 29$ ). No significant difference was found between the use of instruction

in isolation and in combination. In total, Instruction yielded larger effects than all other interventions types.

***Direct Instruction.*** Seven studies evaluated Direct Instruction ( $n = 79$ ). More of those comparisons investigated the use of this intervention in isolation ( $n = 43$ ) than in combination ( $n = 36$ ). No significant difference was found between Direct Instruction used in isolation and used in combination. In total, Direct Instruction yielded larger effects than Curricular Modification and Choice.

***Peer Tutoring.*** Six studies evaluated Peer Tutoring, yielding 65 comparisons. Thirty-two comparisons investigated the use of this intervention in isolation and 33 investigated its use in combination with another intervention. Peer Tutoring used in isolation resulted in a significantly larger effect than when it was used in combination. In total, Peer Tutoring yielded larger effects than Curricular Modification and Choice.

***Punishment.*** Five studies evaluated Punishment (36 comparisons), all of which investigated its use in combination with another intervention. Punishment used in combination yielded larger effects than Choice.

***Function-Based Interventions.*** Four studies evaluated Function-Based Interventions (21 comparisons). Twice as many comparisons investigated the use of this intervention in isolation ( $n = 14$ ) than in combination with another intervention ( $n = 7$ ). No significant difference was found between the use of Function-Based Interventions used in isolation and used in combination (See Table 11).

***Social Skill Instruction.*** Four studies evaluated Social Skill Instruction that yielded 35 comparisons, all of which involved Social Skill Instruction in isolation. Social Skill Instruction yielded larger effects than Choice.

***Peer Modeling.*** Three studies evaluated Peer Modeling. There were 19 comparisons, all of which investigated its use in isolation. Peer Modeling yielded ESs that were equivalent to or larger than all other interventions.

***Curricular Modification.*** Two studies evaluated Curricular Modification, which yielded 30 comparisons. All investigated Curricular Modification used in isolation, and yielded ESs that were equivalent to or less than all other interventions.

***Computer Assisted Instruction.*** One study evaluated Computer Assisted Instruction; it contained 33 comparisons. Twenty-seven comparisons investigated the use of this intervention in isolation and 6 investigated its use in combination with another intervention. No significant difference was found being the use of Computer Assisted Instruction used in isolation and used in combination. Computer Assisted Instruction yielded ESs that were larger than Choice.

***Choice.*** One study evaluated the use of Choice, yielding 15 comparisons, all of which investigated it used in isolation. Choice in isolation yielded effects equivalent or smaller than all other interventions.

**Analysis by ED characteristic.** As discussed earlier, students identified as qualifying for special education under the ED category have one or more of the characteristics found in the federal definition. In order to investigate the extent to which ED characteristics moderate

the effect of interventions, the preceding procedure was repeated for each ED characteristic. The analyses below investigate the effects of interventions, used exclusively or in combination with other interventions, employed separately with students described as having each characteristic of ED plus the two features of Schizophrenia-like and Socially Maladjusted.

*Interventions for students with the (A) Inability to Learn characteristic.* Presented in Table 12 are the effects of interventions, used exclusively or in combination with other interventions, employed with students described as having the (A) Inability to Learn. Refer to Table 7 and recall that overall, interventions used with students with this characteristic produced an Overall PAND of 86.4%, and Overall RPhi coefficient of 0.720 [95% CI = 0.703, 0.735].

The earlier analysis showed that some interventions employed exclusively produced larger effects than the same interventions employed in combination. This same pattern is evident for some interventions used with students with characteristic (A) Inability to Learn.

*Self-Management/Monitoring.* Self-Management/Monitoring was evaluated in 153 comparisons involving students with characteristic (A). More of those comparisons investigated the use of this intervention in isolation ( $n = 115$ ) than in combination with another intervention type ( $n = 38$ ). For students with (A), Self-Management/Monitoring used in isolation resulted in a significantly larger effect than when it was used in combination. Self-Management/Monitoring interventions in total yielded larger effects than Positive Reinforcement, Curricular Modifications and Choice, and smaller effects than Instruction.

*Positive Reinforcement.* Positive Reinforcement was evaluated in 84 comparisons. More of those comparisons investigated the use of this intervention in combination ( $n = 66$ ) than in isolation ( $n = 18$ ). Positive Reinforcement used in isolation resulted in a significantly smaller effect than when it was used in combination. In total, Positive Reinforcement yielded smaller effects than Instruction, Direct Instruction, and Function-Based Interventions.

*Instruction.* Instruction was evaluated in 96 comparisons. More of those comparisons investigated the use of this intervention in isolation ( $n = 76$ ) than in combination ( $n = 20$ ). Instruction used in combination resulted in larger effects than when used in isolation with students with the Inability to Learn. In total, Instruction yielded larger effects than Direct Instruction, Peer Tutoring, Curricular Modification, and Choice.

*Direct Instruction.* Direct Instruction was studied in 39 comparisons. More of those comparisons investigated the use of this intervention in isolation ( $n = 29$ ) than in combination ( $n = 10$ ). No significant difference was found being the use of Direct Instruction used in isolation and used in combination. In total, Direct Instruction yielded larger effects than Positive Reinforcement.

*Peer Tutoring.* Peer Tutoring was evaluated in 28 comparisons. Twenty-two comparisons investigated the use of this intervention in isolation and 6 investigated its use in combination with another intervention. Peer Tutoring used in isolation resulted in a significantly larger effect than when it was used in combination. In total, Peer Tutoring yielded smaller effects than Instruction.

*Function-Based Interventions.* Function-Based Interventions were investigated in 17 comparisons. Most comparisons investigated the use of this intervention in isolation ( $n = 12$ ) than in combination with another intervention ( $n = 5$ ). No significant difference was found being the use of Function-Based Interventions used in isolation and used in combination. In total, Function-Based Interventions yielded larger effects than Positive Reinforcement.

*Social Skill Instruction.* Social Skill Instruction was evaluated in 6 comparisons, all of which were evaluated in isolation. Social Skill Instruction was not significantly different than any other intervention used with students with characteristic (A).

*Curricular Modification.* Curricular Modification was evaluated in 30 comparisons, all of which were investigated when used in isolation. Curricular Modification resulted in were not significantly different than any other intervention.

*Computer Assisted Instruction.* Computer Assisted Instruction was investigated in 10 comparisons. Six comparisons investigated the use of this intervention in isolation and four investigated its use in combination with another intervention. No significant difference was found being the use of Computer Assisted Instruction used in isolation and used in combination. In total, Computer Assisted Instruction produced a larger ES than Positive Reinforcement, and Peer Tutoring,

*Choice.* One study evaluated the use of Choice; it contained 15 comparisons, all of which were investigated used in isolation. Choice resulted in lower effects than Self-Management/Monitoring, Instruction, Direct Instruction, and Computer Assisted Instruction when used with students with the (A) Inability to Learn characteristic.

*Punishment and Peer Modeling.* No studies investigated the use of Punishment or Peer Modeling with students with characteristic (A).

*Interventions implemented with students with (B) Relationship Problems.* Taken together all intervention types used with students described as having (B) Relationship Problems produced an Overall PAND of 79.3%, and Overall RPhi coefficient of 0.581 (95% CI = 0.535, 0.623; see Table 7). Table 13 presents the effects of interventions, used exclusively or in combination with other interventions, employed with students with this characteristic.

*Self-Management/Monitoring.* Self-Management/Monitoring was evaluated with students with (B) Relationship Problems in 20 comparisons. More of those comparisons investigated this intervention in isolation ( $n = 11$ ) than in combination with another intervention type ( $n = 9$ ). Self-Management/Monitoring in isolation resulted in a significantly larger effect than when in combination. In total, Self-Management/Monitoring interventions were not statistically different than other interventions used with this group.

*Positive Reinforcement.* Positive Reinforcement was evaluated in 54 comparisons. More of those comparisons investigated the use of this intervention in combination ( $n = 38$ ) than in isolation ( $n = 16$ ). Positive Reinforcement in isolation resulted in a significantly smaller effect than in combination. In total, Positive Reinforcement produced statistically lower effects than Peer Tutoring interventions used with this group of students.

*Direct Instruction.* Direct Instruction was studied in 10 comparisons. More of those comparisons investigated the use of this intervention implemented in isolation ( $n = 6$ ) than in



combination ( $n = 4$ ). All of those were implemented in combination with Peer Tutoring. No significant difference was found being the use of Direct Instruction in isolation and in combination. Nor were any significant differences found between Direct Instruction in total and any other intervention used with students with (B) Relationship Problems.

*Peer Tutoring.* Peer Tutoring was evaluated in 10 comparisons. Six comparisons investigated this intervention in isolation and 4 investigated it in combination with Direct Instruction (as described above). Peer Tutoring in isolation resulted in a significantly larger effect than when in combination. In total, Peer Tutoring yielded significantly larger effects than Positive Reinforcement.

*Punishment.* Punishment was evaluated in 28 comparisons, all of which investigated its use in combination with other interventions. Punishment did not produce significantly different effect than any other intervention type with students with (B) Relationship Problems.

*Social Skill Instruction.* Social Skill Instruction was evaluated in 13 comparisons, all of which were evaluated in isolation. Social Skill Instruction was not significantly different than any other intervention used with student with characteristic (B).

*Curricular Modification.* Curricular Modification was evaluated in 2 comparisons, both of which were investigated it in isolation. Curricular Modification resulted in larger effects than Positive Reinforcement and Direct Instruction, although this difference is based on only a very few comparisons.

*Instruction, Function-Based Interventions, Peer Modeling, Computer Assisted Instruction, and Choice.* No comparisons evaluated Instruction, Function-based interventions, Peer Modeling, Computer assisted instruction, or Choice with students with characteristic (B).

***Interventions implemented with students with (C) Inappropriate Behavior.***

Presented in Table 14 are the effects of interventions, used exclusively or in combination with other interventions, employed with students described as displaying (C) Inappropriate Behavior. When grouped together, all interventions used with students with this characteristic produced an Overall PAND of 86.7% and Overall RPhi coefficient of 0.734 [95% CI = 0.718, 0.748] (see Table 7).

*Self-Management/Monitoring.* Self-Management/Monitoring was evaluated ( $n = 154$ ) with students with characteristic (C) Inappropriate Behavior. More of those comparisons investigated this intervention in isolation ( $n = 109$ ) than in combination with another intervention type ( $n = 45$ ). Self-Management/Monitoring in isolation resulted in a significantly larger effect than in combination. When used with students with characteristic (C), Self-Management/Monitoring interventions in total yielded larger effects than Choice, and smaller effects than Instruction, Punishment, and Peer Modeling.

*Positive Reinforcement.* Positive Reinforcement was evaluated in 114 comparisons. More of those comparisons investigated this intervention in combination ( $n = 70$ ) than in isolation ( $n = 44$ ). Positive Reinforcement in isolation resulted in a significantly larger effect

than combination. In total, Positive Reinforcement yielded larger effect than Choice, smaller effects than Instruction, and Peer Modeling.

*Instruction.* Instruction was evaluated through 42 comparisons. More of those comparisons investigated this intervention in isolation ( $n = 31$ ) than in combination ( $n = 11$ ). No significant differences were found between Instruction in combination or in isolation with students with (C) Inappropriate Behavior. In total, Instruction yielded larger effects than Self-Management/Monitoring, Positive Reinforcement, Direct Instruction, Peer Tutoring, Social Skill Instruction, Curricular Modifications, and Choice.

*Direct Instruction.* Direct Instruction was studied in 51 comparisons. More of those comparisons investigated this intervention in isolation ( $n = 28$ ) than in combination ( $n = 23$ ). No significant difference was found being the use of Direct Instruction in isolation and in combination. In total, Direct Instruction yielded larger effects than Choice, and smaller effects than Instruction, and Peer Modeling.

*Peer Tutoring.* Peer Tutoring was evaluated in 35 comparisons. Twelve comparisons investigated this intervention in isolation and 23 investigated in combination with another intervention. No significant differences were found between using Peer Tutoring in isolation and in combination with other interventions. In total, Peer Tutoring yielded larger effects than Choice when used with students with characteristic (C).

*Punishment.* Punishment was evaluated in 24 comparisons, all investigated in combination with another intervention. In total, Punishment yielded larger effects than Choice when used with students with characteristic (C).

*Function-Based Interventions.* Function-Based Interventions were investigated in 8 comparisons. More comparisons investigated this intervention in isolation ( $n = 6$ ) than in combination with another intervention ( $n = 2$ ). No significant difference was found between Function-Based Interventions isolation and combination. In total, Function-Based Interventions produced effects that were significantly larger than Choice.

*Social Skill Instruction.* Social Skill Instruction was evaluated in 30 comparisons, all of which were evaluated in isolation. Social Skill Instruction produced effects that were significantly larger than choice with students with characteristic (C).

*Peer Modeling.* Peer Modeling was evaluated in 18 comparisons, all in isolation. Peer Modeling produced effects that were significantly larger than Self-Management/Monitoring, Positive Reinforcement, Peer Tutoring, Social Skill Instruction, Curricular Modification and Choice with students with Inappropriate Behavior.

*Curricular Modification.* Curricular Modification was evaluated in 12 comparisons, all in isolation. Curricular Modification produced effects that were significantly lower than Instruction and Peer Modeling.

*Choice.* Choice was evaluated in 15 comparisons, all of which were investigated in isolation. Choice resulted in significantly lower effects than every intervention other than Curricular Modification.

*Computer Assisted Instruction.* Computer Assisted Instruction was not investigated with students with (C) Inappropriate Behavior.

***Interventions implemented with students with exhibiting (D) Unhappiness or Depression.*** The effects of interventions used exclusively or in combination with other interventions, employed with students described as exhibiting (D) Unhappiness or Depression are presented in Table 15. Recall Table 7 that presented data on all interventions used with students with this characteristic and produced an Overall PAND of 82.8%, and an Overall RPhi coefficient of 0.628 (95% CI = 0.583, 0.670).

***Self-Management/Monitoring.*** Self-Management/Monitoring was evaluated in 13 comparisons. More of those comparisons investigated the use of this intervention in isolation ( $n = 10$ ) than in combination with another intervention type ( $n = 3$ ). Self-Management/Monitoring in isolation resulted in a significantly smaller effect than when in combination. When used with students with (D) Unhappiness or Depression, Self-Management/Monitoring interventions in total yielded smaller effects than Positive Reinforcement, Instruction, and Curricular Modification.

***Positive Reinforcement.*** Positive Reinforcement was evaluated in 15 comparisons. More of those comparisons investigated this intervention in combination ( $n = 11$ ) than in isolation ( $n = 4$ ). No significant differences were found between Positive Reinforcement in combination or in isolation with students with (D) Unhappiness or Depression. In total, Positive Reinforcement yielded larger effects than Self-Management/Monitoring and Choice.

***Instruction.*** Instruction was evaluated through 10 comparisons. Half those comparisons investigated this intervention in isolation ( $n = 5$ ) and the other half investigated it in combination ( $n = 5$ ). No significant differences were found between Instruction in

combination or in isolation with students with characteristic (D). In total, Instruction yielded larger effects than Self-Management/Monitoring, and Choice.

*Direct Instruction.* Direct Instruction was studied in 15 comparisons. More of those comparisons investigated this intervention in combination ( $n = 9$ ) than in isolation ( $n = 6$ ). No significant difference was found between Direct Instruction in isolation and in combination. In total, Direct Instruction in total did not result in effects that were significantly different than any other intervention.

*Peer Tutoring.* Peer Tutoring was evaluated in 16 comparisons. Four comparisons investigated the use of this intervention in isolation and 12 investigated it in combination with another intervention. No significant differences were found between using Peer Tutoring and in combination with other interventions. Further no significant differences were found between Peer Tutoring in total and any other intervention.

*Social Skill Instruction.* Social Skill Instruction was evaluated in 15 comparisons, all of which were evaluated in isolation. In total, Social Skill Instruction did not produce effects that were significantly different than any other intervention used with students with characteristic (D).

*Curricular Modification.* Curricular Modification was evaluated in 3 comparisons, all of which were investigated used in isolation. Curricular Modification did not produce effects that were significantly different than any other intervention type.

*Choice.* Choice was evaluated in 10 comparisons, all of which were investigated in isolation. Choice resulted in lower effects than Positive Reinforcement and Instruction.

*Function-Based Interventions, Punishment, Peer Modeling, and Computer Assisted Instruction.* No SCR studies were found to investigate the use of Function-based interventions, Punishment, Peer Modeling, and Computer Assisted Instruction with students with (D) Unhappiness or Depression characteristic.

*Interventions implemented with students with (E) Physical Symptoms or Fears.* As a whole, interventions used with students with Physical Symptoms or Fears produced an Overall PAND of 79.4%, and Overall RPhi coefficient of 0.550 [95% CI = 0.477, 0.616] (Table 7). The effects of interventions used exclusively or in combination with other interventions, employed with students with this characteristic are presented in Table 16.

*Self-Management/Monitoring.* Self-Management/Monitoring was evaluated in 10 comparisons. All of those comparisons investigated the use of this intervention in isolation ( $n = 10$ ). Self-Management/Monitoring produced lower effect than Positive Reinforcement when used with students with characteristic (E).

*Positive Reinforcement.* Positive Reinforcement was evaluated in 8 comparisons. More of those comparisons investigated this intervention in combination ( $n = 6$ ) than in isolation ( $n = 2$ ). No significant differences were found between using Positive Reinforcement used in isolation or in combination. In total, Positive Reinforcement in total yielded larger effects than Self-Management/Monitoring, and Direct Instruction.

*Instruction.* Instruction was evaluated through 6 comparisons. Half of those comparisons investigated this intervention in isolation ( $n = 3$ ) and the remaining investigated it in combination ( $n = 3$ ). No significant differences were found between the Instruction in

combination or in isolation with students with (E) Physical Symptoms or Fears. In total, Instruction did not differ significantly from other interventions used with this group.

*Direct Instruction.* Direct Instruction was studied in 5 comparisons. Two comparisons investigated this intervention in isolation and three in combination. No significant difference was found between Direct Instruction in isolation and in combination. In total, Direct Instruction yielded smaller effects than Positive Reinforcement.

*Peer Tutoring.* Peer Tutoring was evaluated in 6 comparisons, all of which were investigated in combination with another intervention. No significant differences were found between Peer Tutoring and other interventions.

*Choice.* Choice was evaluated in 5 comparisons, all of which were investigated in isolation. No significant differences were found between Choice and any other intervention.

*Punishment, Function-Based Interventions, Social Skill Instruction, Peer Modeling, Curricular Modification, and Computer Assisted Instruction.* No SCR studies were found that investigated Punishment, Function-Based Interventions, Social Skill Instruction, Peer Modeling, Curricular Modification, and Computer Assisted Instruction with students with (E) Physical Symptoms or Fears.

***Interventions implemented with students with Schizophrenia-Like Features.*** Recall Table 7 that presents overall interventions used with students with Schizophrenia-Like Features produced an Overall PAND of 91.4%, and an Overall RPhi coefficient of 0.828 [95% CI = 0.737, 0.891]. The effects of interventions used exclusively or in combination with other interventions, employed with students with this characteristic are presented in



Table 17. No significant differences were found between interventions used with students exhibiting Schizophrenia-Like Features.

*Self-Management/Monitoring.* Self-Management/Monitoring was evaluated in two comparisons, both of which were investigated in combination with another intervention type ( $n = 2$ ).

*Positive Reinforcement.* Positive Reinforcement was evaluated in five comparisons, all of which were investigated this intervention in combination ( $n = 5$ ).

*Instruction.* Instruction was evaluated through 4 comparisons. Half of those comparisons investigated this intervention in isolation ( $n = 2$ ), with the other half in combination ( $n = 2$ ). No significant differences were found between Instruction in combination or in isolation with students with Schizophrenia-Like Features.

*Direct Instruction and Peer Tutoring.* Direct Instruction was studied in 5 comparisons, all of which were investigated combination with Peer Tutoring ( $n = 5$ ).

*Social Skill Instruction.* Social Skill Instruction was evaluated in 2 comparisons, all of which were evaluated in isolation.

*Punishment, Function-Based Interventions, Peer Modeling, Curricular Modification, Computer Assisted Instruction, and Choice.* Punishment, Function-Based Interventions, Peer Modeling, Curricular Modification, Computer Assisted Instruction, and Choice were not investigated with students with Schizophrenia-Like Features.

***Interventions implemented with Socially Maladjusted students.*** The effects of interventions exclusively or in combination with other interventions, employed with Socially

Maladjusted students as are presented in Table 18. Recall that taken together, interventions used with students with this characteristic produced an Overall PAND of 83.9%, and an Overall RPhi coefficient of 0.678 [95% CI = 0.624, 0.726] (see Table 7).

*Self-Management/Monitoring.* Self-Management/Monitoring was evaluated in 10 comparisons. All of those comparisons investigated this intervention in isolation. When used with students described as Socially Maladjusted, Self-Management/Monitoring interventions yielded smaller effects than Positive Reinforcement, Direct Instruction, and Peer Tutoring.

*Positive Reinforcement.* Positive Reinforcement was evaluated in 4 comparisons, all of which investigated this intervention in isolation. Positive Reinforcement in isolation resulted in a significantly larger effect than Self-Management/Monitoring, Direct Instruction, Peer Tutoring, and Social Skill Instruction.

*Direct Instruction.* Direct Instruction was studied in 10 comparisons, all of which were implemented in combination with another intervention. Direct Instruction yielded larger effects than Self-Monitoring/Management and Social Skill Instruction.

*Peer Tutoring.* Peer Tutoring was evaluated in 20 comparisons. Half of the comparisons investigated the use of this intervention in isolation ( $n = 10$ ) and the other half investigated its use in combination with another intervention ( $n = 10$ ). No significant differences were found between using Peer Tutoring in and in combination with other interventions. In total, Peer Tutoring yielded larger effects than Self-Management/Monitoring, and Social Skill Instruction.

*Social Skill Instruction.* Social Skill Instruction was evaluated in 5 comparisons, all of which were evaluated in isolation. Social Skill Instruction produced effects that were significantly smaller than Positive Reinforcement, Direct Instruction, and Peer Tutoring with students described as Socially Maladjusted.

*Instruction, Punishment, Function-Based Interventions, Peer Modeling, Curricular Modification, Computer Assisted Instruction, and Choice.* Instruction, Punishment, Function-Based Interventions, Peer Modeling, Curricular Modification, Computer Assisted Instruction, and Choice were not investigated in with Socially Maladjusted students.

### **Summary**

In this chapter, results were presented for this study's three research questions. On the basis of Overall PAND analyses of the primary single case research on interventions used with students with ED, the effect size is large.

Overall PAND analyses of the single case research on the effect of interventions varied by ED characteristic, although all produced moderate effect sizes. Students described as exhibiting (B) Relationship Problems, (D) Unhappiness or Depression, or (E) Physical Symptoms or Fears demonstrated lower effects, compared to those with the (A) Inability to Learn, (C) Inappropriate Behavior, Schizophrenia-Like Features, and were Socially Maladjusted. These results are for all interventions, and are not broken down by type of intervention.

Further, present Overall PAND analyses of intervention types revealed varied effects for students with different ED characteristics. For each ED characteristic, different

interventions were found to have varying effects. Moreover, some interventions were shown to have larger effects when used in combination, whereas others were shown to have larger effects when used in isolation. Based on the present sample of SCR studies, particular ED characteristics appear to respond differently to different types of interventions.

## CHAPTER 5

As the field of education becomes increasingly guided by empirical evidence of effectiveness, and the corpus of evidence in the form of scientific research on education grows, education leaders and practitioners need to make use of methods to aggregate and synthesize that evidence. Teachers, administrators, and educator trainers must be enabled to use this evidence to make data-based decisions to meet the needs of students, including students with ED. One of the tenets inherent in special education is that some students require teaching and management that is different or “special” in their education. The extending logic therefore suggests that different intervention approaches may be more or less effective with particular populations of students. Present evidence (see Chapter 4) indicates this may be true for our most challenging students, those with ED.

The present meta-analysis not only shows that different interventions can and do impact student outcomes for students with ED, but those outcomes vary by student characteristics as defined in the federal definition of ED. Understanding this variation will encourage teachers to make data-based decisions about appropriate interventions for students and increase the likelihood of significant improvement in both academic achievement and behavior.

In this chapter, results are discussed as related to the three research goals of this study. First, I discuss findings as to whether and to what extent interventions improve outcomes for students with ED. Next, I address how ED characteristics appear to moderate the effectiveness of interventions. Third, I discuss which interventions are most effective for

students with different ED characteristics. In addition, other observations and implications are also discussed.

### **To What Extent do Interventions Improve Outcomes for Students with ED?**

Chapter 4 *Results* revealed significant effect size statistics for interventions in general. These results indicate that interventions for students with ED, as evaluated in published SCR studies, resulted in a large positive effect; this applied not to any particular intervention, but to performing an intervention as opposed to continuing in baseline (ordinary) circumstances. Given the significant and severe educational, behavioral and social needs of students with ED (Kauffman 2005), this finding should motivate researchers to continue applying, creating, and studying interventions for students with ED.

**Outcomes by setting.** That 525 of the 636 comparisons by setting (see Table 9) occurred in “more restrictive” education environments suggests that emphasis is needed on SCR studies that take place in Resource and General Education settings. There was no correlation found between publication date and setting. This fails to support the potential explanation that the field’s increasing preference for less restrictive settings influences the settings studied. One possible explanation may be that it is easier to study participants in captive settings (e.g., Self-Contained classrooms and separate settings).

Students with ED placed in less restrictive environments (i.e., general education, resource), were linked to larger effects than those in more restrictive environments (i.e., self-contained, separate settings). One possible explanation is that students in less restrictive environments had behaviors that were not as severe as those in more restrictive settings. This

would indicate that those with less severe behaviors were able to make larger gains than those with more significant needs.

**Outcomes by student sex.** While there has been much discussion about the disproportionate sex representation in ED (Donovan & Cross, 2002; Wagner, et. al., 2005), there was no statistical difference between effect sizes for males and females (when sex was reported). It is difficult to explain why participants for whom sex was not reported demonstrated significantly larger outcomes than either males or females.

**Outcomes by student race.** The race of the participants was not reported for most of the participants included in this meta-analysis. For those studies in which race was reported, Caucasian and Hispanic participants demonstrated significantly larger effects than African American participants. It is conceivable that significantly lower effects for African American students may indicate that these students do not respond as well to the interventions examined.

Alternatively, this finding may also be the result of increased awareness about this disproportionate representation of African American students in the special education referral and identification process (Donovan & Cross, 2002; Wagner, et. al., 2005). Local education agencies may feel pressure to be careful about “over identifying” African American students for ED, and only refer and place those students with the greatest and most severe need. Such students may be particularly resistant to interventions. As seen in the differential outcomes of students by educational placement (a potential proxy for severity of need), the African American students in the included studies may have produced lower effects because they

may have greater level of need. In either case, more research is needed to clarify the cause of these lower effects.

Very few studies reported the sex or the race of the main intervention agent (e.g., teacher, researcher). It is possible that this is an important consideration. For example, perhaps the power of some interventions is a joint function of the race of intervention agent and the race of the student. Therefore I recommend that investigators include the sex and race of the interventionist in their reports.

**Academic outcomes.** Students with ED, on average, exhibit significant academic deficits (Wagner, Kutash, Duchnowski, Epstein, & Sumi, 2005). These students often experience academic difficulties such as low or failing grades, high retention rates, and an increased likelihood of dropping out (Reschley & Christenson, 2006). Several interventions are appropriate and effective in improving academic outcomes, including Self-Management/Monitoring, Instruction, Direct Instruction, Peer-Tutoring, Curricular Modifications, and Computer-Assisted Instruction (see Chapter 4).

Yet there is evidence that some of these interventions are more likely to improve academic outcomes than others. In rank order (not necessarily statistically significant) from greatest ES to least: Computer Assisted Instruction in combination, Computer Assisted Instruction in isolation, Instruction in combination, Instruction in isolation, Positive Reinforcement in isolation, Direct Instruction in isolation, Direct Instruction in combination, Self-Management/ Monitoring in isolation, Positive Reinforcement in combination, Peer Tutoring in isolation, Self-Management/ Monitoring in combination, Curricular



Modifications in isolation, and Peer Tutoring in combination. These results could have implications for how teachers select interventions to be used with students with ED.

Teachers may be able to identify which intervention to implement first, by considering which is most likely to produce the best outcome.

Further, some interventions are more powerful when implemented in isolation than in combination with others. Self-Management/ Monitoring, Positive Reinforcement, and Peer Tutoring used in isolation produced a significantly larger effect than when used in combination. These data suggest that for these interventions, more is not better. Combining multiple interventions does not appear to have an additive effect.

**Behavioral outcomes.** As a group, students with ED have serious, multiple, and complex behavioral/social problems that begin early in life (Wagner, Kutash, Duchnowski, Epstein, & Sumi, 2005). Several interventions are appropriate and effective in improving behavioral outcomes, including Self-Management/Monitoring, Positive Reinforcement, Punishment, Function-Based Interventions, Social Skill Instruction, Peer Modeling, and Choice.

As with academic outcomes, some of these interventions are more likely to improve behavioral outcomes than others. In rank order from most confidence of effectiveness to least: Peer Modeling in isolation, Positive Reinforcement in isolation, Punishment in combination, Function-Based Interventions in isolation, Punishment in combination, Self-Management/ Monitoring in isolation, Punishment in combination, Social Skill Instruction in

isolation, Choice in isolation, Positive Reinforcement in combination, Self-Management/ Monitoring in combination, and Function-Based Interventions in combination.

Again some interventions are more powerful when implemented in isolation than in combinations with others. Self-Management/ Monitoring, Positive Reinforcement, and Function Based Interventions used in isolation produced a significantly larger effect than when used in combination. Any additional investment in combining multiple interventions does not appear to improve student outcomes. However, this points to the need for new research that directly compares the use in isolation versus in combination, to clarify these research leads.

### **Do ED Characteristics Moderate The Effectiveness Of Interventions?**

Table 7 summarizes the disaggregated findings regarding the five characteristics of ED plus *Socially Maladjusted* and *Schizophrenia*, and shows that characteristics do in fact moderate intervention effectiveness. In rank order of students with ED characteristics with whom showed most confidence of effectiveness to least: Schizophrenia-Like Features, (A) Inability to Learn, (C) Inappropriate Behavior, Socially Maladjusted, (D) Unhappiness or Depression, (E) Physical Symptoms or Fears, and (B) Relationship Problems. Certainly many students exhibit more than one of these characteristics, and this is true of the participants in this meta-analysis. Additional research is needed to understand the effect of interventions used with students with only one of these characteristics, as well as the effect of specific combinations of characteristics.

Many research reports on interventions used with students with ED did not contain detailed information about which of the five ED characteristics participants exhibited. However, the distribution of those with specific ED characteristics reported was heavily skewed. Over 400 participants included in this meta-analysis were reported as exhibiting (A) Inability to Learn and (C) Inappropriate Behavior. By contrast only 39 participants were reported as Socially Maladjusted, 32 exhibiting Physical Symptoms or Fears, and 13 with Schizophrenia-Like Features. Researchers appear to be most interested in students who struggle academically and those with aggressive, disruptive behavior problems. Conversely, they appear to be less interested in students with other ED characteristics. One may wonder if there are fewer students with the other characteristics, but research indicates that all five characteristics of ED are abundant (e.g., Cullinan & Epstein, 2001) These lower numbers of participants limit what we know about how they respond to interventions general, and to specific interventions (discussed next).

### **Which Interventions Are Most Effective For Students With Different ED Characteristics?**

Beyond the data that reflect students with different characteristics responding differently to different interventions, the present meta-analysis also found that ED characteristics differentially moderate the effectiveness of interventions. Teachers can use these results to select interventions that are most likely to produce the largest gains for students with specific characteristics. That is, making intervention decisions based on students with ED as a group may be a disservice to students with some specific

characteristics. Below, I present several examples of the interaction of characteristics of ED by kind of intervention.

For example, Positive Reinforcement used in isolation produced a significantly larger effect size over all students with ED, than when used specifically with students with characteristic (A) Inability to Learn, or students with (B) Relationship Problems.

In another example, Peer Tutoring used in combination with another intervention produced a significantly larger effect when used specifically with students with (C) Inappropriate Behavior and Socially Maladjusted students, than with the total group of students with ED. So Peer Tutoring may be even more effective for students with (C) and Socially Maladjusted students than would be expected for the total group of students with ED.

Other examples can be cited. Positive reinforcement used in isolation produced significantly larger effects with students with characteristic (C) Inappropriate Behavior than those with characteristic (A) Inability to Learn. Similarly, Self-Management/Monitoring implemented in isolation produced significantly larger effects for students with characteristic (A) Inability to Learn, than those with characteristic (D) Unhappiness or Depression. In yet another example, Positive Reinforcement in isolation produced a large effect for students with characteristic (E) Physical Symptoms or Fears, and a moderate effect for students with characteristic (A) Inability to Learn.

The general point is that according to the present analysis, several widely-used educational interventions are differentially effective, depending on which characteristics of

ED is present in the behavior of the student receiving the intervention. Especially if confirmed by appropriate future research. This information could affect which interventions are selected and used, which may improve the educational practice for students with ED.

As mentioned above, the vast majority of the participants for whom ED characteristics were reported, exhibited (A) Inability to Learn and (C) Inappropriate Behavior. Therefore more research was available on interventions used with participants with these characteristics. Yet, little or no SCR was located that investigated the use of some interventions with these groups. No effects could be calculated for Punishment and Peer Modeling used with students with characteristic (A), nor Computer Assisted Instruction used with students with characteristic (C). Perhaps these interventions have not been deemed appropriate for students with these characteristics, and consequently not studied. Alternatively, these are areas in need of additional research.

### **Research Implications**

**Inconsistent term usage.** Inconsistent and varying definitions and terms have been used to describe students with ED and EBD. Inconsistent term usage makes it more difficult to compare studies. While many in the field use the term ED to refer to the population of students that qualify for special education services under the federal definition, the broader term EBD includes any student who has emotional or behavioral problems. While the participants in this meta-analysis had to qualify for special education services under the category of ED, several researchers use the more general term of EBD even when clearly describing students that qualify for special education services. Perhaps some researchers

modify the term to be more acceptable to particular journals, which may use the more general term in their title (e.g., *Behavioral Disorders, Journal of Emotional and Behavioral Disorders*). Another explanation may be that the term “disorder” sounds more socially acceptable than “disturbance”. While this study was not designed to determine the cause of this term confusion, it does reflect that this inconsistent term use does exist in the professional literature.

These terms used inconsistently make it difficult for readers to generalize research findings. This inconsistency of terms made this meta-analysis more challenging, and is likely to confuse practitioners trying to interpret research and select appropriate empirically proven interventions. Perhaps this contributes to the research-to-practice gap. It would greatly benefit the field of ED to find consensus on terms and their related definitions.

**Analysis by SCR design.** Differential effects may be artifacts of the SCR design. Multiple baseline across probes design ( $k=7$ , PAND = 97.7%, RPhi = 0.955 [0.922, 0.975]) produced effects that were significantly larger than any other design (See table 9). Perhaps studies in which probes are created to measure change are more sensitive to that change than assessments that measure behavior change in other ways. Conversely, complex ABC designs produced outcomes that were significantly smaller than any other design. It is possible that those designs are less sensitive to change, or were applied predominately to outcome variables that are highly resistant to change. Further research is needed with other populations to determine if this difference in effect is the consequence of the design rather than the effect of the application of an intervention.

*Analysis by research fidelity and quality.* There were more comparisons included in this meta-analysis originating from research reports that did not describe intervention fidelity measures, than did. Further, those reports that did report intervention implementation fidelity measures produced larger effects than those that did not. This difference in effect size may point to the value of measuring and documenting fidelity of intervention implementation. Perhaps the act of recording implementation fidelity ensures that interventions are implemented more consistently and accurately, there by resulting in better outcomes for students (O'Donnell, 2008). An alternative explanation could be that those studies that conducted fidelity checks were correlated with student behaviors that were easier to improve, thus were more likely to be published in peer-reviewed journals, and consequently included in this meta-analysis. The value of intervention fidelity in school interventions with ED is an important research question.

*Analysis by What Works Clearinghouse criteria.* The criteria set by What Works Clearinghouse have been critiqued as being so stringent that few extant studies meet the criteria. Yet more studies than were expected met the evidence standards or met those standards with reservations. Recall that WWC criteria require IOA checks on at least 20% of observations in each phase. Many research reports describe their IOA procedures with an statement of the percent of observations for the whole study on which IOA was assessed. This statement is often followed up with another statement indicating that IOA was assessed at least once in each phase. Very few studies report the percentage of IOA checks per phase. The potentially unintended consequence of the WWC criteria favors those studies with five

or less data points per phase. Since relatively few research reports explicitly described their IOA procedures with such detail, many studies met this criterion on the basis of either short phase lengths or 100% IOA checks. Studies with 5 or less data points per phase whose author's reported IOA checks on at least 20% of all observations, "with at least one observation per phase" also met the criteria. One IOA check of five (or fewer) observations met or exceeded the 20% minimum. Studies that have more than 5 data points per phase that used the same verbiage did not meet the criteria. This phenomenon gives more weight to studies with shorter phases over those with longer phases when IOA procedures are not described in explicit detail.

Also, studies that conducted IOA checks on all observations also met this criterion. Again, those studies that did meet the WWC criteria produced outcomes that had significantly larger effect sizes than those that did not. This may be indicative of the same phenomenon described above with intervention implementation fidelity reporting. Well-designed studies with significant effect sizes are more likely to be published and included in this meta-analysis.

### **Limitations**

This meta-analysis possesses several limitations, including those that are the result of the search procedure, statistical analyses, and study report quality.

**Search procedure.** Although a thorough search was conducted, it is possible that some relevant articles were not identified and included in the analysis. Because of the high number of possible descriptors that could be used to search the electronic databases for



articles to be included in this review, it is possible that I omitted of some descriptors that may have resulted in failure to identify some articles. Second, the criteria used to select articles may have affected the findings. For example, some articles may be excluded because only those studies conducted in the United States and Canada were included. Third, articles that did not provide data separately for students with ED were excluded, which may have altered the results. Finally, all reviewed studies were those published in peer-reviewed journals. This leaves open the possibility that unpublished “file drawer” research produced results counter to those in published reports, leading to results different than those presented in this study.

**Statistical analysis.** No single best approach to summarizing one SCR study and aggregating SCR has been identified, and all current approaches have limitations. While the combined approach used in this meta-analysis was selected to mitigate those limitations, some still exist.

For instance, when there is no data overlap between baseline and treatment phases, PAND awards a 100% score (along with a correspondingly high RPhi), regardless of the distance between data in those phases (i.e., level or magnitude of change from baseline to treatment phases). This gives an advantage to studies with observations that yield minimal, but consistent, improvement to avoid any overlap of data between phases. In other words, PAND does not take into account magnitude of difference between phases. Another limitation to PAND is that it does not control for baseline trends. When a positive trend (trend in the desired direction) is present in the baseline phase(s), some phenomena outside

the experimental control may also be influencing the outcome. Therefore, some of the effect in the treatment phase may be attributed to those phenomena. PAND does not remove that effect from the calculation. Finally and perhaps most importantly, a large effect size alone does not imply that change was due to the intervention.

The strength of the research design must also be considered. The convenience and ease of which ES can be summarized and interpreted may entice some to forget to consider the quality of research design (i.e., reduced threats to internal validity) in deciding whether or not a change is due to the application of an intervention. Some threats to internal validity include: ambiguous temporal precedence (i.e., lack of clarity about which variable occurred first); history (i.e., events occurring concurrently with the intervention); maturation (i.e., naturally occurring changes over time); statistical regression toward the mean (i.e., when participants are selected on the basis of their extreme scores, their scores on subsequent measures may be less extreme); attrition (i.e., loss of respondents); testing (i.e., exposure to a test can affect scores on subsequent exposures to that test); instrumentation (i.e., the conditions or nature of a measure might change over time); and additive and interactive effects (i.e., the impact of a threat can be added to that of another threat or may be moderated by levels of another threat) which confuse the intervention effect. For some, publication may serve as a proxy measurement of research design quality. Others may question the inconsistent standard held by editors and peer-reviewers, and seek another scheme to evaluate study design, such as what has been put forward by WWC as well as others (Horner, et. al., 2003).

**Characteristics.** Students with ED often present multiple characteristics. However, this analysis only investigated the outcomes per single characteristic. That is, each student was assigned to every ED characteristic described in that student's study, but characteristics of ED were examined one at a time. This study did not investigate combinations of characteristics.

Some interventions were not investigated with participants with some ED characteristics (e.g, Schizophrenia-Like Features, Socially Maladjusted, (E) Physical Symptoms or Fears). Additional research is needed to understand the effect of interventions used with these less-studied populations.

**Study quality.** The outcomes of this meta-analysis are limited by the quality, completeness, and accuracy of the primary studies from which it drew. Recall that not all studies reported the independent variables used in this analysis. The outcomes reported here may be impacted by the fact that some variables (e.g., participant sex, race) were not consistently reported.

**Representative population.** The participants included in this meta-analysis may not have been representative of the national population of students with ED. A higher percentage of participants in this study (86%) were male than in the national population (80%). Race was not reported for over half of the participants included in this study. Of those for whom race was reported, 48% were Caucasian, 45% African American and 8% were Hispanic, which contrasts with the national race proportions of 57%, 27%, and 12% respectively (Wagner, Kutash, Duchnowski, Epstein, & Sumi, 2005). Due to incomplete

reporting, we cannot be certain if the population included in this analysis is representative of the total population of students with ED. Therefore, caution should be taken in generalizing these results to the larger population of students with ED.

### **Summary**

This meta-analysis identified several findings that may be used to as basis for further investigation. Published SCR studies of interventions used with students with ED indicate that interventions can change behavior and academic outcomes. Students with specific ED characteristics may respond to intervention differently. Characteristics of ED are associated with variable effects for different intervention types. Some interventions may be more effective for students with some characteristics than others. Further, the act of measuring implementation fidelity may also improve outcomes for students. Certainly, many of these findings need to be investigated through additional research.

## REFERENCES

- Aaore, L., & Nelson J.R. (2000). A comparative analysis of teachers' Caucasian parents' and Hispanic parents' views of problematic school survival behaviors. *Education and Treatment of Children, 23*(3), 314.
- Ager, C. L., & Cole, C. L. (1991). A review of cognitive-behavioral interventions for children and adolescents with behavioral disorders. *Behavioral Disorders, 16*(4), 276-87.
- \*Alber, S. R., Anderson, L. L., Martin, C. M., & Moore, K. J. (2005). Teaching elementary students with behavior disorders to recruit positive teacher attention: Effects on math proficiency. *Journal of Early and Intensive Behavior Intervention, 2*(1), 59-73.
- Ang, R. P., & Hughes, J. N. (2002). Differential benefits of skills training with antisocial youth based on group composition: A meta-analytic investigation. *School Psychology Review, 31*(2), 164-185.
- \*Anthony, L., Wolery, M., Werts, M. G., Caldwell, N. K., & Snyder, E. D. (1996). Effects of daily probing on acquisition of instructive feedback responses. *Journal of Behavioral Education, 6*(2), 111-133. Doi:10.1007/BF02110228
- Armstrong, K. H., Dedrick, R. F., & Greenbaum, P. E. (2003). Factors associated with community adjustment of young adults with serious emotional disturbance: A longitudinal analysis. *Journal of Emotional and Behavioral Disorders, 11*, 66-75.

- Baer, R. A., & Nietzel, M. T. (1991). Cognitive and behavioral treatment of impulsivity in children: A meta-analytic review of the outcome literature. *Journal of Clinical Child Psychology, 20*(4), 400-412.
- \*Barton-Arwood, S. M., Wehby, J. H., & Falk, K. B. (2005). Reading instruction for elementary-age students with emotional and behavioral disorders: academic and behavioral outcomes. *Exceptional Children, 72*(1), 7-27.
- Beard-Jordan, K., & Sugai, G. (2004). First Step to Success: An early intervention for elementary children at risk for antisocial behavior. *Behavioral Disorders, 29*, 396-409.
- \*Beare, P., Torgerson, C., & Creviston, C. (2008). Increasing verbal behavior of a student who is selectively mute. *Journal of Emotional and Behavioral Disorders, 16*(4), 248-255. Doi:10.1177/1063426608317356
- Beelman A., Pfungsten, U., & Losel, F. (1994) Effects of training social competence in children: A meta-analysis of recent evaluation studies. *Journal of Clinical Child Psychology, 23*(3), 260.
- Beelmann, A., & Pfungsten, U. (1994). Effects of training social competence in children: A meta-analysis of recent evaluation studies. *Journal of Clinical Child Psychology, 23*(3), 260-283.
- \*Bell, K., Young, K., Blair, M., & Nelson, R. (1990). Facilitating mainstreaming of students with behavioral disorders using classwide peer tutoring. *School Psychology Review, 19*(4), 564-573.

- Benner, G. J., Nelson, J. R., Ralston, N. C., & Mooney, P. (2010). A meta-analysis of the effects of reading instruction on the reading skills of students with or at risk of behavioral disorders. *Behavioral Disorders, 35*(2), 86-102.
- \*Billingsley, G., Scheuermann, B., & Webber, J. (2009). A comparison of three instructional methods for teaching math skills to secondary students with emotional/behavioral disorders. *Behavioral Disorders, 35*(1), 4-18.
- Blackorby, J., & Wagner, M. (1996). Longitudinal postschool outcomes of youth with disabilities: Findings from the National Longitudinal Transition Study. *Exceptional Children, 62*, 399-414.
- \*Blankenship, T. L., Ayres, K. M., & Langone, J. (2005). Effects of computer-based cognitive mapping on reading comprehension for students with emotional behavior disorders. *Journal of Special Education Technology, 20*(2), 15-23.
- Borenstein, M, Hedges, L, Higgins J & Rothstein, HR (2005). *Comprehensive Meta-Analysis, Version 2*. Englewood, NJ: Biostat, Inc.
- Borenstein, M., & Rothstein, H. (2005). *Comprehensive Meta-Analysis*. Englewood, NJ: Biostat. I believe the one below is the correct one. It has all creators.
- Bower, G. H. & Cohen, P. R. (1982). Emotional Influences on Memory and Thinking: Data and Theory. In S. Fiske & M. Clark (eds), *Affect and Cognition*. Hillsdale, NJ. Erlbaum.

Bower, G. H., Clark, M. C., Lesgold, A. M. & Winzenz, D. (1969). Hierarchical Retrieval Schemes in Recall of Categorized Word Lists. *Journal of Verbal Learning and Verbal Behaviour*, 8, p. 323 – 343.

\*Bowman-Perrott, L. J., Greenwood, C. R., & Tapia, Y. (2007). The efficacy of CWPT used in secondary alternative school classrooms with small teacher/pupil ratios and students with emotional and behavioral disorders. *Education & Treatment Of Children*, 30(3), 65-87. Doi:10.1353/etc.2007.0014

Bradley, R., Doolittle, J., & Bartolotta, R. (2008). Building on the data and adding to the discussion: The experiences and outcomes of students with emotional disturbance. *Journal Emotional and Behavioral Disorders*, 17(1), 4-23.

\*Brasch, T. L., Williams, R., & McLaughlin, T. F. (2008). The Effects of a Direct Instruction Flashcard System on Multiplication Fact Mastery by Two High School Students with ADHD and ODD. *Child & Family Behavior Therapy*, 30(1), 51-59.

Burns, M. K., & Wagner, D. (2008). Determining an effective intervention within a brief experimental analysis for reading: A meta-analytic review. *School Psychology Review*, 37(1), 126-136.

Burns, M. K., Coddling, R. S., Boice, C. H., & Lukito, G. G. (2010). Meta-analysis of acquisition and fluency math interventions with instructional and frustration level skills: Evidence for a skill-by-treatment interaction. *School Psychology Review*, 39(1), 69-83.



- \*Callicott, K. J., & Park, H. (2003). Effects of self-talk on academic engagement and academic responding. *Behavioral Disorders, 29*(1), 48-64.
- Campbell, D.T., & Stanley, J.C. (1966). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.
- \*Cancio, E. J., West, R. P., & Young, K. R. (2004). Improving mathematics homework completion and accuracy of students with EBD through self-management and parent participation. *Journal of Emotional and Behavioral Disorders, 12*(1), 9–22.
- Carnine, D., & Thomas B. Fordham Foundation, W. C. (2000). *Why Education Experts Resist Effective Practices (And What It Would Take To Make Education More Like Medicine)*.
- \*Carr, S. C., & Punzo, R. P. (1993). The effects of self-monitoring of academic accuracy and productivity on the performance of students with behavioral disorders. *Behavioral Disorders, 18*(4), 241-250.
- Carver, R. (1978). The Case Against Statistical Significance Testing. *Harvard Educational Review, 48*(3), 378-399.
- Christenson, C. L. & Thurlow, M. L. (2004). School dropouts: Prevention considerations, interventions, and challenges. *Current Directions in Psychological Science, 13* (10). 36-39. DOI: 10.1111/j.0963-7214.2004.01301010.x
- Christenson, S. L., & Thurlow, M. L. (2004). Keeping kids in school: Efficacy of Check & Connect for dropout prevention. *NASP Communiqué, 32*(6), 37–40.

- \*Cieslar, W., McLaughlin, T. F., & Derby, K. (2008). Effects of the copy, cover, and compare procedure on the math and spelling performance of a high school student with behavioral disorder: A case report. *Preventing School Failure, 52*(4), 45-51.
- Clarke, S., Dunlap, G., & Stichter, J. (2002). A descriptive analysis of intervention research in emotional and behavioral disorders from 1980 through 1999. *Behavior Modification, 26*(5), 659-683. Doi:10.1177/014544502236656
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2<sup>nd</sup> ed.).
- Coleman, M., & Vaughn, S. (2000). Reading interventions for students with emotional/behavioral disorders. *Behavioral Disorders, 25*, 93-104.
- Cook, C.R., Gresham, F.M., Kern, L., Barreras, R.B., Thornton, S., & Crews, S.D. (2008). Social skills training for secondary students with emotional and/or behavioral disorders: A review and analysis of the meta-analytic literature. *Journal of Emotional and Behavioral Disorders*. Doi: 10.117/1063426608314541.
- Cooper, H., & Hedges, L.V. (1994). *The Handbook of Research Synthesis*. New York, Russell Sage.
- Coutinho, M.J. (1986). Reading achievement of students identified as behaviorally disordered at the secondary level. *Behavioral Disorders, 11*, 200-207.
- Cullinan, D. (2004). Research issues in classification and definition of emotional and behavioral disorders. In M. Quinn, S. Mathur, & R.B. Rutherford, Jr. (Eds.), *Handbook of research in behavioral disorders* (pp. 32-53). New York: Guilford.

- Cullinan, D., & Epstein, M. H. (2001). Comorbidity among Students with Emotional Disturbance. *Behavioral Disorders, 26*(3), 200-13.
- Cullinan, D., & Sabornie, E. J. (2004). Characteristics of emotional disturbance in middle and high school students. *Journal of Emotional and Behavioral Disorders, 12*(3), 157-167.
- Cullinan, D., Evans, C., Epstein, M. H., & Ryser, G. (2003). Characteristics of emotional disturbance of elementary school students. *Behavioral Disorders, 28*, 94–110.
- \*Daly, E., Garbacz, S., Olson, S. C., Persampieri, M., & Ni, H. (2006). Improving oral reading fluency by influencing students' choice of instructional procedures: An experimental analysis with two students with behavioral disorders. *Behavioral Interventions, 21*(1), 13-30. Doi:10.1002/bin.208
- DSM-IV-TR mental disorders: diagnosis, etiology, and treatment (2004). Published: Chichester, Hoboken, NJ: J. Wiley, American Psychological Association, 2004
- DSM-IVTR; American Psychological Association, 2004.
- Donovan, M. S., & Cross, C. T. (Eds.). (2002). *Minority students in special and gifted education*. Washington, DC: National Academy Press.
- \*Ducharme, J. M., & Harris, K. E. (2005). Errorless Embedding For Children With On-Task And Conduct Difficulties: Rapport-Based, Success-Focused Intervention In The Classroom. *Behavior Therapy, 36*(3), 213-222. Doi:10.1016/S0005-7894(05)80070-X
- \*Dunlap, G., White, R., Vera, A., Wilson, D., & Panacek, L. (1996). The effects of multi-component, assessment-based curricular modifications on the classroom behavior of

- children with emotional and behavioral disorders. *Journal Of Behavioral Education*, 6(4), 481-500. Doi:10.1007/BF02110518
- Dunst, C. J., Hamby, D. W., & Trivette, C. M. (2004). Guidelines for calculating effect sizes for practice-based research synthesis. *Centerscope*, 3(1), 1-10. Retrieved from <http://www.researchtopractice.info/productCenterscope.php>.
- \*DuPaul, G. J. & McGoey, K. E. (1997). Mainstreaming students with behavior disorders: The use of classroom peers as facilitators of generalization. *School Psychology Review*, 26(4), 634-650.
- Durlak, J. A., Fuhrman, T., & Lampman, C. (1991). Effectiveness of cognitive-behavior therapy for maladapting children: A meta-analysis. *Psychological Bulletin*, 110(2), 204-214.
- Education of All Handicapped Children Act (PL.94-142)
- Education Sciences Reform Act (ESRA, 2002)
- Epstein, M. H., & Cullinan, D. (1992). Assessment practices used in programs for adolescents with behavior disorders. *Preventing School Failure*, 36(3), 20-25.
- Epstein, M. H., & Cullinan, D. (2009). *Scales for assessing emotional disturbance* (2nd ed.). Austin, TX: Pro-Ed.
- Epstein, M. H., Nelson, J. R., Trout, A. L., & Mooney, P. (2004). Achievement and emotional disturbance: Academic status and intervention research. In M. H. Epstein, K. Kutash, & A. J. Duchnowski (Eds.), *Outcomes for children with emotional and*

- behavioral disorders and their families: Programs and evaluation of best practices* (Vol. 2). Austin TX: Pro-Ed.
- Evans, J., Harden, A., & Thomas, J. (2004). What are effective strategies to support pupils with emotional and behavioural difficulties in mainstream primary schools? Findings from a systematic review of research. *Journal of Research in Special Educational Needs*, 4(1), 2-16. Doi:10.1111/J.1471-3802.2004.00015.x
- \*Falk, K. B., & Wehby, J.H. (2001). The effects of peer-assisted learning strategies on the beginning reading skills of young children with emotional or behavioral disorders. *Behavioral Disorders*, 26, 344-359.
- Federal Register*. (2006). Assistance to states for the education of children with disabilities and preschool grants for children with disabilities; Final Rule 34. CFR Parts 300 and 301, 71 Fed. Reg. 46540 (August 14, 2006).
- \*Ferguson, A., Ashbaugh, R., O'Reilly, S., & McLaughlin, T. F. (2004). Using prompt training and reinforcement to reduce transition times in a transitional kindergarten program for students with severe behavior disorders. *Child & Family Behavior Therapy*, 26(1), 17-24. Doi:10.1300/J019v26n01\_02
- Forness, S. R. and Kavale, K.A. (1994). The Balkanization of special education: Proliferation of categories for 'new' behavioral disorders. *Education & Treatment of Children*, 17 (3). 215-231.

Forness, S.R., & Knitzer, J. (1992). A new proposed definition and terminology to replace “serious emotional disturbance” in Individuals with Disabilities Education Act.

*School Psychology Review, 21, 12-20.*

\*Foxx, R. M., Kyle, M. S., Faw, G. D., & Bittle, R. G. (1989). Problem-solving skills training: Social validation and generalization. *Behavioral Residential Treatment, 4(4)*, 269-288. Doi:10.1002/bin.2360040402

\*Freeman & Dexter-Mazza (2004). Looks like this one should be 2005. Looked at Dexter-Mazza’s vitae, and it only has a 2005 pub.

Gage, N. A., Lewis, T. J., & Stichter, J. P. (2012). Functional behavioral assessment-based interventions for students with or at risk for emotional and/or behavioral disorders in school: A hierarchical linear modeling meta-analysis. *Behavioral Disorders, 37(2)*, 55-77.

Gast, D. L. (2010). *Single Subject Research Methodology in Behavioral Sciences*. New York, NY: Routledge.

Gersten, R., Fuch, L.S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M.S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Council for Exceptional Children, 71 (2)*, 149-164.

Ghafoori, B., & Tracz, S. M. (2001). Effectiveness of cognitive-behavioral therapy in reducing classroom disruptive behaviors: A meta-analysis.

<http://www.eric.ed.gov/PDFS/ED457182.pdf>.

Glass, G. V. (1997). *Education Policy Analysis Archives, 5*, 1-22 1997.

- Glass, G. V. (Ed.) (1976). *Evaluation Studies Review Annual*, Vol. 1. Beverly Hills: Sage Publications.
- Glass, G. V., McGaw, B. & Smith, M. L. (1981). *Meta-analysis in Social Research*. New York, NY: Sage.
- Gould, R. A., Safren, S. A., Washington, D. O., & Otto, M. W. (2004). A meta-analytic review of cognitive-behavioral treatments. In R. G. Heimberg, C. L. Turk, & D. S. Mennin (Eds.), *Generalized Anxiety Disorder: Advances In Research And Practice*. (pp. 248-264). New York, NY US: Guilford Press.
- Greenbaum, P. E., Dedrick, R. F., Friedman, R. M., Kutash, K., Brown, E. C., Lardierh, S. P., & et al. (1996). National Adolescent And Child Treatment Study (NACTS): Outcomes for children with serious emotional and behavioral disturbance. *Journal of Emotional and Behavioral Disorders*, 4, 130-146.
- Gresham, F. M. (2005). Treatment integrity and therapeutic change: Commentary on Perepletchikova and Kazdin. *Clinical Psychology: Science and Practice*, 12(4), 391-394. Doi:10.1093/clipsy/
- Gresham, F. M. (2007). Response to intervention and emotional and behavioral disorders: Best practices in assessment for intervention. *Assessment for Effective Intervention*, 32, 214-222.
- Griffith, A. K., Duppong Hurley, K., & Hagaman, J. L. (2009). Treatment integrity of literacy interventions for students with emotional and/or behavioral disorders: A

- review of literature. *Remedial and Special Education*, 30(4), 245-255.  
Doi:10.1177/0741932508321013
- Griffith, A., Hurley, K., & Hagaman, L. (2009). Treatment of literacy interventions for students with emotional and/or behavioral disorders. *Remedial and Special Education*, 30, 245-255.
- \*Gulchak, D. J. (2008). Using a mobile handheld computer to teach a student with an emotional and behavioral disorder to self-monitor attention. *Education & Treatment of Children*, 31(4), 567-581. Doi:10.1353/etc.0.0028
- Gulchak, D. J., & Lopes, J. A. (2007). Interventions for students with behavioral disorders: An international literature review. *Behavioral Disorders*, 32(4), 267-281.
- \*Gunter, P. L., Miller, K. A., & Venn, M. L. (2003). A case study of the effects of self-graphing reading performance data for a girl identified with emotional/behavioral disorders. *Preventing School Failure*, 48(1), 28-31.
- \*Hale, A. D., Skinner, C. H., Winn, B. D., Oliver, R., Allin, J. D., & Molloy, C. M. (2005). An investigation of listening and listening-while-reading accommodations on reading comprehension levels and rates in students with emotional disorders. *Psychology In The Schools*, 42(1), 39-51. Doi:10.1002/pits.20027
- Hallahan, D., Keller, C., & Ball, D. (1986). A comparison of prevalence rate variability from state to state for each of the categories in special education. *Remedial and Special Education*, 7, 8-14.



- \*Hansen, S. D., & Lignugaris/Kraft, B. (2005). Effects of a dependent group contingency on the verbal interactions of middle school students with emotional disturbance. *Behavioral Disorders, 30*(2), 170-184.
- \*Hawkins, R. O., & Axelrod, M. I. (2008). Increasing the on-task homework behavior of youth with behavior disorders using functional behavioral assessment. *Behavior Modification, 32*(6), 840-859. Doi:10.1177/0145445508318846
- Heckaman, K., Conroy, M., Fox, J. J., & Chait, A. (2000). Functional assessment-based intervention research on students with or at risk for emotional and behavioral disorders. *Behavioral Disorders, 25*, 196-210.
- Hedges L.V.&Olkin I. (1985). *Statistical methods for meta-analysis*. New York: Academic Press.
- Hedges, L. V., & Olkin, I. (1984). Nonparametric estimators of effect size in meta-analysis. *Psychological Bulletin, 96*(3), 573-580.
- Hoagwood, K. E., Olin, S., Kerker, B. D., Kratochwill, T. R., Crowe, M., & Saka, N. (2007). Empirically based school interventions targeted at academic and mental health functioning. *Journal of Emotional and Behavioral Disorders, 15*(2), 66-92. Doi:10.1177/10634266070150020301
- Holinger, P.C. (1987). *Violent deaths in the United States: An epidemiologic study of suicide, homicide, and accidents*. New York, NY: Guilford Press.

- Horner, R., Carr, E., Halle, J., McGee, G., Odom, S., & Wolery, M. (2003). The use of single-case research to identify evidenced-based practice in special education. *Exceptional Children, 71*(2), 165-179.
- \*Hughes, C.A., Deshler, D.D., Ruhl, K.L. & Schumaker (1993) Test-taking strategy instruction for adolescents with emotional and behavioral disorders. *Journal of Emotional and Behavioral Disorders, 1*, 189-198.
- Jaffee, S.R., Harrington, H., Cohen, P., Moffitt, T.E. (2005). Cumulative prevalence of psychiatric disorder in youths. *Journal of the American Academy of Adolescent Psychiatry, 44*(5): 406-407.
- Jensen, W. R., Clark, E., Kircher, J. C. & Kristijansson, S. D. (2007). Statistical reform: Evidence-based practice, meta-analyses, and single subject designs. *Psychology in the Schools, 44*, 483-493.
- \*Jolivette, K., Wehby, J.H., Canale, J., & Massey, N.G. (2001). Effects of choice making opportunities on the behavior of students with emotional and behavioral disorders. *Behavioral Disorders, 26*, 131-145.
- Joseph, L. M. & Eveleigh, E. (2011). A review of the effects of self-monitoring on the reading performance of students with disabilities. *Journal of Special Education, 45*, 43-53.
- Kauffman, J. M., Mock, D. R., & Simpson, R. L. (2007). Problems related to underservice of students with emotional or behavioral disorders. *Behavioral Disorders, 33*, 43-57.

- Kauffman, J. M. (2005). Waving to Ray Charles: Missing the meaning of disability. *Phi Delta Kappa*, 86, 520-521, 524.
- Kauffman, J.M. & Landrum, T.J. (2009). *Characteristics of emotional and behavioral disorders of children and youth (9<sup>th</sup> ed.)*. Upper Saddle River, NJ: Prentice Hall.
- Kauffman, J. M., Mock, D. R., & Simpson, R. L. (2007). Problems related to underservice of students with emotional or behavioral disorders. *Behavioral Disorders*, 33, 43-57.
- Kavale, K. A., (1979). Mainstreaming: The genesis of an idea. *The Exceptional Child*, 26, 3-21.
- Kazdin, A. E. (2012). A comprehensive guide to the rich and rigorous research methods of psychological science. *Psyccritiques*, 57(21), Doi:10.1037/a0028394
- Kazdin, A. E. (2011a). Evidence-based research: Advances, limitations, and next steps. *American Psychologist*, 66, 21-37.
- Kazdin, A. E. (2011b). *Single-case research designs: Methods for clinical and applied settings (2<sup>nd</sup> edition)*. New York: Oxford University Press.
- \*Kern-Dunlap, L., Dunlap, G., Clarke, S., & Childs, K. E. (1992). Effects of a videotape feedback package on the peer interactions of children with serious behavioral and emotional challenges. *Journal of Applied Behavior Analysis*, 25(2), 355-364.  
Doi:10.1901/jaba.1992.25-355
- \*Kern, L., Childs, K. E., Dunlap, G., Clarke, S., & Falk (1994). Using assessment-based curricular intervention to improve the classroom behavior of a student with emotional and behavioral challenges. *Journal of Applied Behavior Analysis*, 27, 7-19.

- \*Kern, L., Delaney, B., Clarke, S., Dunlap, G., & Childs, K. (2001). Improving the classroom behavior of students with emotional and behavioral disorders using individualized curricular modifications. *Journal of Emotional and Behavioral Disorders, 9*(4), 239-247. Doi:10.1177/106342660100900404
- \*Kern, L., Dunlap, G., Childs, K. & Clarke, S. (1994). Student-assisted functional assessment interview. *Diagnostique 19 (2-3)*, 29-39.
- \*Kern, L., Wacker, D. P., Mace, F., & Falk, G. D. (1995). Improving the peer interactions of students with emotional and behavioral disorders through self-evaluation procedures: A component analysis and group application. *Journal of Applied Behavior Analysis, 28*(1), 47-59. Doi:10.1901/jaba.1995.28-47
- Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational And Psychological Measurement, 56*(5), 746-59.
- Kline, R. B. (2004). *Beyond significance testing: Reforming data analysis methods in behavioral research*. Washington, DC: American Psychological Association.
- Kratochwill, T. R., & Stoiber, K.C., (2002). Evidence-based interventions in school psychology: Conceptual foundations of the Procedural and Coding manual of Division 16 and the Society for the Study of School Psychology Task Force. *School Psychology Quarterly, 17*, 341-389.
- Kratochwill, T. R., Hitchcock, J. J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D., & Shadish, What Works Clearinghouse, (2010). Single-Case Designs Technical Documentation. *What Works Clearinghouse*,

- Kutash, K., & Duchnowski, A. J. (2004). The mental health needs of youth with emotional and behavioral disabilities placed in special education programs in urban schools. *Journal of Child and Family Studies, 13*(2), 235-248.
- Lakin, M. (1984). Trying to bridge psychoanalysis and group psychotherapy. *Psyocritiques, 29*(7), 568-569. Doi:10.1037/023031
- Landrum, T. J., & Lloyd, J. W. (1992). Generalization in social behavior research with children and youth who have emotional or behavioral disorders. *Behavior Modification, 16*(4), 593-616. Doi:10.1177/01454455920164009
- Lane, K. L. (2004). Academic instruction and tutoring interventions for students with emotional/behavioral disorders: 1990 to present. In R. B. Rutherford, M. M. Quinn, & S. R. Mathur (Eds.), *Handbook Of Research In Emotional And Behavioral Disorders* (pp. 317–326). New York: Guilford Press.
- Lane, K. L., Carter, E. W., Pierson, P. R., & Glaeser, B. C. (2006). Academic, social, and behavioral characteristics of high school students with emotional disturbances or learning disabilities. *Journal of Emotional and Behavioral Disorders, 14*(2), 108-107.
- Lane, K., Eisner, S. L., Kretzer, J., Bruhn, A. L., Crnabori, M., Funke, L., & Casey, A. (2009). Outcomes of functional assessment-based interventions for students with and at risk for emotional and behavioral disorders in a job-share setting. *Education & Treatment of Children, 32*(4), 573-604. Doi:10.1353/etc.0.0073

- \*Lawson, T., & Greer, R. (2006). Teaching the function of writing to middle school students with academic delays. *Journal of Early and Intensive Behavior Intervention*, 3(1), 151-170.
- Lewis, T. J., Hudson, S., Richter, M., & Johnson, N. (2004). Scientifically supported practices in emotional and behavioral disorders: a proposed approach and brief review of current practices. *Behavioral Disorders*, 29(3), 247-259.
- Lipsey, M. W., & Wilson, D. B. (2001). Practical meta-analysis. Applied Social Research Methods Series (Vol. 49). Thousand Oaks, CA: Sage Publications, Inc.
- \*Lloyd, M. E., & Hilliard, A. M. (1989). Accuracy of self-recording as a function of repeated experience with different self-control contingencies. *Child & Family Behavior Therapy*, 11(2), 1-14. Doi:10.1300/J019v11n02\_01
- \*Locke, W. R., & Fuchs, L. S. (1995). Effects of peer-mediated reading instruction on the on-task behavior and social interaction of children with behavior disorders. *Journal of Emotional and Behavioral Disorders*, 3(2), 92-99.
- Lonigan, C. J., Elbert, J. C., & Johnson, S. B. (1998). Empirically supported psychosocial interventions for children: An overview. *Journal of Clinical Child Psychology*, 27, 138-145.
- Losel, F. & Beelman, A. (2003). Effects of child skills training on preventing antisocial behavior: A systematic review of randomized evaluations. *Annals of the American Academy of Political and Social Science*, 587, 84-109.

- Ma, H. H. (2006). An alternative method for quantitative synthesis of single-subject research: Percentage of data points exceeding the median. *Behavior Modification, 30*, 598-617.
- Maag, J. W. (2006) Social skills training for youth with emotional and behavioral disorders: A review of reviews. *Behavioral Disorders, 32*, 4-17.
- Magee, M., Quinn, M., Kavale, K. A., Mathur, S. R., Rutherford, R. R., & Forness, S. R. (1999). A meta-analysis of social skill interventions for students with emotional or behavioral disorders. *Journal of Emotional and Behavioral Disorders, 7*(1), 54-64.  
Doi:10.1177/106342669900700106
- Maggin, D. M., Swaminathan, H., Rogers, H. J., O’Keeffe, B. V., Sugai, G., & Horner, R. H. (2011). A generalized least squares regression approach for computing effect sizes in single-case research: Application examples. *Journal of School Psychology, 49*(3), 301-321. Doi:10.1016/j.jsp.2011.03.004
- \*Maher, C. A. (1986). Direct replication of a cross-age tutoring program involving handicapped adolescents and children. *School Psychology Review, 15*, 100-118.
- Mathur, S. R., Kavale, K. A., Quinn, M., Forness, S. R., & Rutherford, R. r. (1998). Social skills interventions with students with emotional and behavioral problems: A quantitative synthesis of single-subject research. *Behavioral Disorders, 23*(3), 193-201.

- Mathur, S. R. & Rutherford, R. B., (1991). Peer-mediated interventions promoting social skills of children and youth with behavioral disorders. *Education and Treatment of Children, (14(3), 227-242.*
- Matyas, T.A., & Greenwood, K.M. (1990). Visual analysis of single-case time series: Effects of variability, serial dependence, and magnitude of intervention effects. *Journal of Applied Behavior Analysis, 23(3), 341-51.*
- May, K. (2003). A note on the use of confidence intervals. *Understanding Statistics, 2(2), 133-135.*
- \*McCurdy, B. L., Cundari, L., & Lentz, F. (1990). Enhancing instructional efficiency: An examination of time delay and the opportunity to observe instruction. *Education & Treatment of Children, 13(3), 226-238.*
- \*McCurdy, B. L., & Shapiro, E. S. (1988). Self-observation and the reduction of inappropriate classroom behavior. *Journal of School Psychology, 26(4), 371-378.*  
Doi:10.1016/0022-4405(88)90044-1
- McIntosh, R., Vaughn, S., & Zargoza, N. (1991). A review of social interventions for students with learning disabilities. *Journal of Learning Disabilities, 24, 451-458.*
- \*McKenny, A. & Dattilo, J. (2001). Effects of an intervention within a sport context of the prosocial behavior and antisocial behavior of adolescents with disruptive behavior disorders. *Therapeutic Recreation Journal, 35(2), 123-140.*
- \*McLaughlin, T. F. (1992). Effects of written feedback on reading with behaviorally disordered students. *Journal of Education Research, 85, 312-316.*



Merikanga, K.R., He, J.P., Brody, D., Fisher, P.W., Bourdon, K., Koretz, D.S., (2010).

Prevalence and treatment of mental disorders among US children in the 2001-2004 NHANES. *Pediatrics*, *125*(1): 75-81.

Merrell, K. W., & Walker, H. M. (2004). Deconstructing a definition: Socially Maladjusted versus emotional disturbance and moving the EBD field forward. *Psychology In The Schools*, *41*(8), 899-910. Doi:10.1002/pits.20046

\*Miller, D. N., & Cole, C. L. (1998). Effects of social skills training on an adolescent with comorbid conduct disorder and depression. *Child & Family Behavior Therapy*, *20*(1), 35-53.

Mitchell, C., & Hartmann, D.P. (1981). A cautionary note on the use of omega squared to evaluate the effectiveness of behavioral treatments. *Behavioral Assessment*, *3*, 93-100.

Mooney, P., Ryan, J. B., Uhing, B. M., Reid, R., & Epstein, M. H. (2005). A review of self-management interventions targeting academic outcomes for students with emotional and behavioral disorders. *Journal of Behavioral Education*, *14*(3), 203-221.

Doi:10.1007/s10864-005-6298-1

\*Moore, R. J., Cartledge, G. & Heckaman, (1995). The effects of social skills instruction and self-instruction and self-monitoring on game-related behaviors of adolescents with emotional or behavior disorders. *Behavioral Disorders*, *20*, 253-266.

\*Mulcahy, C. A., Krezmien, M. P., Leone, P. E., Houchins, D. E., & Baltodano, H. (2008). Lessons learned: Barriers and solutions for conducting reading investigations in

- juvenile corrections settings. *Reading & Writing Quarterly: Overcoming Learning Difficulties*, 24(2), 239-252. Doi:10.1080/10573560701808635
- \*Musser, E. H., Bray, M. A., Kehle, T. J., & Jenson, W. R. (2001). Reducing disruptive behaviors in students with serious emotional disturbance. *School Psychology Review*, 30(2), 294-304.
- National Council on Disability (2004). *Improving Educational Outcomes for Students with Disabilities*. Retrieved from: <http://www.ncd.gov/publications/2004/Mar172004>.
- Nelson, C., Rutherford, R. B., Center, D. B., & Walker, H. M. (1991). Do public schools have an obligation to serve troubled children and youth?. *Exceptional Children*, 57(5), 406-415.
- Nelson, C., Rutherford, R., Center, D. & Walker, H. (1991). Do public schools have an obligation to serve troubled children and youth? *Exceptional Children*, 57, 406-415.
- Nelson, J. R., Benner, G. J., Lane, K., & Smith, B. W. (2004). Academic achievement of K-12 students with emotional and behavioral disorders. *Exceptional Children*, 71(1), 59-73.
- Nelson, J., Lane, K. L., Benner, G. J., & Kim, O. (2011). A best evidence synthesis of literacy instruction on the social adjustment of students with or at-risk for behavior disorders. *Education & Treatment Of Children*, 34(1), 141-162.  
Doi:10.1353/etc.2011.0000

Nelson, J., Smith, D. J., Young, R. K., & Dodd, J. M. (1991). A review of self-management outcome research conducted with students who exhibit behavioral disorders.

*Behavioral Disorders, 16(3)*, 169-179.

Nelson, J. R., Smith, D. J., Young, R. K. & Dodd, J. M. (1991). A review of self-management outcomes research conducted with students who exhibit behavioral disorders. *Behavioral Disorders, 16(3)*, 169-179.

\*Nelson, W. M., & Behler, J. J. (1989). Cognitive impulsivity training: The effects of peer teaching. *Journal of Behavior Therapy and Experimental Psychiatry, 20(4)*, 303-309.

Doi:10.1016/0005-7916(89)90061-X

Newman, L., Wagner, M., Cameto, R., Knokey, A., & National Center for Special Education Research (ED). (2009). The post-high school outcomes of youth with disabilities up to 4 years after high school: a report from The National Longitudinal Transition Study-2 (NLTS2). NCSER 2009-3017. National Center for Special Education Research. Retrieved from [http://www.lib.ncsu.edu/cgi-](http://www.lib.ncsu.edu/cgi-bin/proxy.pl?server=http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED505448&site=ehost-live&scope=site)

[bin/proxy.pl?server=http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED505448&site=ehost-live&scope=site](http://www.lib.ncsu.edu/cgi-bin/proxy.pl?server=http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED505448&site=ehost-live&scope=site)

\*Ninness, H., Ellis, J., Miller, W., & Baker, D. (1995). The effect of a self-management training package on the transfer of aggression control procedures in the absence of supervision. *Behavior Modification, 19(4)*, 464-490.

Doi:10.1177/01454455950194004

- \*Ninness, H., Fuerst, J., & Rutherford, R. (1995). A descriptive analysis of disruptive behavior during pre- and post-unsupervised self-management by students with serious emotional disturbance: A within-study replication. *Journal of Emotional and Behavioral Disorders, 3*(4), 230-240. Doi:10.1177/106342669500300405
- O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. *Review of Educational Research, 78*(1), 33-84.
- O'Shaughnessy, T., Lane, K. L., Gresham, F. M., & Beebe-Frankenberger, M. (2003). Children placed at risk for learning and behavioral difficulties: Implementing a school-wide system of early identification and prevention. *Remedial and Special Education, 24*, 27-35. DOI: 10.1177/074193250302400103
- Odom, S. L. (2009). The tie that binds: Evidence-based implementation, science, and early intervention. *Topics in Early Childhood Special Education, 29*, 53-61.
- Odom, S. L., Brantlinger, E., Gersten, R., Horner, R.H., Thompson, B., & Harris, K.R. (2005). Research in special education: Scientific methods and evidence-based practices. *Exceptional Children, 71*, 137-148.
- \*Ogier, R., & Hornby, G. (1996). Effects of differential reinforcement on the behavior and self-esteem of children with emotional and behavioral disorders. *Journal of Behavioral Education, 6*(4), 501-510. Doi:10.1007/BF02110519
- Olmeda, R. E., & Kauffman, J. M. (2003). Sociocultural considerations in social skills training research with African American students with emotional or behavioral

- disorders. *Journal of Developmental and Physical Disabilities*, 15(2), 101-121.  
Doi:10.1023/A:1022871232435
- Palmaffy, T., (2001). Head Start: The war on poverty goes to school. *Education Next*, 1(2), 10-11.
- Parker, R. I., & Vannest, K. (2009). An improved effect size for single-case research: Nonoverlap of all pairs. *Behavior Therapy*, 40(4), 357-367.  
Doi:10.1016/j.beth.2008.10.006
- Parker, R. I., Hagan-Burke, S., & Vannest, K. (2007). Percentage of all non-overlapping data (PAND): An alternative to PND. *The Journal of Special Education*, 40(4), 194-204.  
Doi:10.1177/00224669070400040101
- Parker, R. I., Vannest, K. J., & Brown, L. (2009). The “Improvement Rate Difference” for single-case research. *Exceptional Children*, 75(2), 135-150.
- Parker, R. I., Vannest, K. J., & Davis, J. L. (2011). Effect size in single-case research: A review of nine nonoverlap techniques. *Behavior Modification*, 35(4), 303-322.  
Doi:10.1177/0145445511399147
- Parker, R. I. & Brossart, D. F. (2003). Evaluating single-case data: A comparison of seven statistical methods. *Behavior Therapy*, 34, 189-211.
- Parker, R. I. & Hagan-Burke, S. (2007). Useful effect sizes interpretations for single case research. *Behavioral Therapy*, 38, 95-105.
- Parker, R. I. & Vannest, K. J. (2009). An improved effect size for single case research: Non-overlap of all pairs (NAP). *Behavior Therapy*, (40), 357-367

- Parker, R. I., Vannest, K. J., Davis, J. L., Sauber, S. B. (2011). Combining nonoverlap and trend for single-case research: Tau-u. *Behavior Therapy, 42*(2), 284-299.
- Parsonson, B. S., & Baer, D. M. (1992). The visual analysis of data, and current research into the stimuli controlling it. In T. R. Kratochwill, J. R. Levin (Eds.), *Single-case research design and analysis: New directions for psychology and education* (pp. 15-40). Hillsdale, NJ England: Lawrence Erlbaum Associates, Inc.
- Patterson, K. B. (2005). Increasing positive outcomes for African American males in special education with the use of guided notes. *The Journal of Negro Education, 74*(4), 311-320.
- Pierce, C. D., Reid, R., & Epstein, M. H. (2004). Teacher-mediated interventions for children with EBD and their academic outcomes. *Remedial and Special Education, 25*, 175-188.
- \*Popkin, J., & Skinner, C. H. (2003). Enhancing academic performance in a classroom serving students with serious emotional disturbance: Interdependent group contingencies with randomly selected components. *School Psychology Review, 32*(2), 282-295.
- \*Prater, M. A., Hogan, S., & Miller, S. R. (1992). Using self-monitoring to improve on-task behavior and academic skills of an adolescent with mild handicaps across special and problem-solving training: A review. *Remedial and Special Education, 14*, 25-37.

- Quinn, M.M., Kavale, K.A., Mathur, S.R., Rutherford, R.B. & Forness, S.R. (1999). A meta-analysis of social skill interventions for students with emotional or behavioral disorders. *Journal of Emotional and Behavioral Disorders*, 7, 54-64.
- \*Rafferty, L. A., & Raimondi, S. L. (2009). Self-monitoring of attention versus self-monitoring of performance: Examining the differential effects among students with emotional disturbance engaged in independent math practice. *Journal of Behavioral Education*, 18(4), 279-299. Doi:10.1007/s10864-009-9092-7
- \*Rasmussen, K. & O'Neill, R. E. (2006). The effects of fixed-time reinforcement schedules on problem behavior of children with emotional and behavioral disorders in a day-treatment classroom setting. *Journal of Applied Behavior Analysis*, 39 (4), 453-457.
- Reddy, L.A., DeThomas, C.A., Newman, E., & Chun, V. (2009). School-based prevention and intervention programs for children with emotional disturbance: A review of treatment components and methodology. *Psychology in the Schools*, 46(2), 132-153.
- Redfield, D.L., & E.W. Rousseau (1981). A meta-analysis of experimental research on teacher questioning behavior. *Review of Education Research*, 51, 237-245.
- Reid, R., Gonzalez, J. E., Nordness, P. D., Trout, A., & Epstein, M. H. (2004). A meta-analysis of the academic status of students with emotional/behavioral disturbance. *Journal of Special Education*, 37, 130-143.
- Reschly, A., & Christenson, S.L. (2006). School completion. In G.G. Bear & K.M. Minke (Eds.), *Children's needs III: Development, prevention, and intervention* (pp.103-113). Bethesda, MD: National Association of School Psychologists.

- Rosenthal, R. (1976). *Experimenter effects in behavioral research* (enlarged ed.). New York: Irvington.
- Rosenthal, R. (1984). *Meta-analytic procedures for social research* (rev. ed.). Newbury Park, CA: Sage.
- Rosenthal, R. (1991). Meta-analysis: A review. *Psychosomatic Medicine*, *53*, 247-271.
- Rosenthal, R. (1994). Parametric measures of effect size. In H. M. Cooper & L. V. Hedges (Eds.), *Handbook of research synthesis* (pp. 231–244). New York: Russell Sage Foundation.
- Rosenthal, R. (1995). Writing meta-analytic reviews. *Psychological Bulletin*, *118*, 183–192.
- Rosnow, R. L., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist*, *44*, 1276–1284.
- Rutherford, R. R., Mathur, S. R., Council for Children with Behavioral, D. s., Teacher Educators for Children with Behavioral, D. s., & Arizona State Univ., T. e. (1992). *Severe Behavior Disorders of Children and Youth. Monograph in Behavioral Disorders, Volume 15*.
- Ryan, J. B., Reid, R., & Epstein, M. H. (2004). Peer-mediated intervention studies on academic achievement for students with emotional and behavioral disorders. *Remedial and Special Education*, *25*(6), 330-342.
- Ryan, J. B., Reid, R., & Epstein, M. H. (2004). Peer-mediated intervention studies on academic achievement for students with ebd: a review. *Remedial And Special Education*, *25*(6), 330-341. Doi:10.1177/07419325040250060101



- Sabornie, E. J., Cullinan, D., Osborne, S. S., & Brock, L. B. (2005). Intellectual, academic and behavioral functioning of students with high-incidence disabilities: A cross-categorical meta-analysis. *Exceptional Children, 72*(1), 47-63. Doi:Article
- \*Salend, S. J., Reynolds, C. J., & Coyle, E. M. (1989). Individualizing the good behavior game across type and frequency of behavior with emotionally disturbed adolescents. *Behavior Modification, 13*(1), 108-126. Doi:10.1177/01454455890131007
- \*Salend, S. J., & Sonnenschein, P. (1989). Validating the effectiveness of a cooperative learning strategy through direct observation. *Journal Of School Psychology, 27*(1), 47-58. Doi:10.1016/0022-4405(89)90030-7
- Schloss, P. J., Schloss, C. N., Wood, C. E., & Kiehl, W. S. (1986). A critical review of social skills research with behaviorally disordered students. *Behavioral Disorders, 12*(1), 1-14.
- Schneider, B. H., & Byrne, B. M. (1984). Predictors of successful transition from Self-Contained special education to regular class settings. *Psychology In The Schools, 21*(3), 375-380. Doi:10.1002/1520-6807(198407)21:3<375::AID-PITS2310210317>3.0.CO;2-Z
- Schneider, N., Goldstein, H., & Parker, R. (2008). Social skills interventions for children with autism: A meta-analytic application of percentage of all nonoverlapping data (PAND). *Evidence-Based Communication Assessment and Intervention, 2*, 152-162.

\*Schoenfeld, N. A., & Mathur, S. R. (2009). Effects of cognitive-behavioral intervention on the school performance of students with emotional or behavioral disorders and anxiety. *Behavioral Disorders, 34*(4), 184-195.

Scruggs, T. E., & Mastropieri, M. A. (1998). Summarizing single subject research: Issues and applications. *Behavior Modification, 22*, 221-242.

Scruggs, T. E., Mastropieri, M. A., & Casto, G. (1987). 'Meta-analysis for single-subject research: When does it clarify, when does it obscure?': Response to Salzberg, Strain, and Baer. *RASE: Remedial & Special Education, 8*(2), 49-52.

Doi:10.1177/074193258700800210

Scruggs, T. E., Mastropieri, M. A., & Casto, G. (1987). 'The quantitative synthesis of single-subject research: Methodology and validation': Reply to Owen White. *RASE: Remedial & Special Education, 8*(2), 40-42. Doi:10.1177/074193258700800208

Scruggs, T. E., Mastropieri, M. A., & Richter, L. (1985). Peer tutoring with behaviorally disordered students: Social and academic benefits. *Behavioral Disorders, 10*(4), 283-294.

Scruggs, T. E., & Mastropieri, M.A. (1994). The utility of the PND statistic: A reply to Allison and Gorman. *Behavior Modification, 22*, 221-242.

Shadish, W. R., & Rindskopf, D. M. (2007). Methods for evidence-based practice: quantitative synthesis of single-subject designs. *New Directions For Evaluation, 113*, 95-109.

- Sinclair, M. F., Christenson, S. L., & Thurlow, M. L. (2005). Promoting school completion of urban secondary youth with emotional or behavioral disabilities. *Exceptional Children, 71*(4), 465-482.
- Singh, N. N., Deitz, D. E., Epstein, M. H., & Singh, J. (1991). Social behavior of students who are seriously emotionally disturbed: A quantitative analysis of intervention studies. *Behavior Modification, 15*(1), 74-94. Doi:10.1177/01454455910151005
- \*Singh, N. N., Lancioni, G. E., Joy, S., Winton, A. W., Sabaawi, M., Wahler, R. G., & Singh, J. (2007). Adolescents with conduct disorder can be mindful of their aggressive behavior. *Journal of Emotional and Behavioral Disorders, 15*(1), 56-63.  
Doi:10.1177/10634266070150010601
- Skeels & Dye, 1939; Kirk, 1958 [as reported by Odom et al., 2005]).
- Skiba, R. & Grizzle, K. (1991). The socially maladjusted exclusion: Issues of definition and assessment. *School Psychology Review, 20*, 577-595.
- Skiba, R., & Casey, A. (1985). Interventions for behaviorally disordered students: A quantitative review and methodological critique. *Behavioral Disorders, 10*(4), 239-252.
- Skiba, R., & Grizzle, K. (1992). Qualifications vs. logic and data: Excluding conduct disorders from the SED definition. *School Psychology Review, 21*(1), 23-28.
- \*Skinner, C. H., Belfiore, P. J., & Pierce, N. (1992). Cover, copy, and compare: Increasing geography accuracy in students with behavior disorders. *School Psychology Review, 21*(1), 73-81.

- \*Skinner, C. H., Hurst, K. L., Teeple, D. F., & Meadows, S. O. (2002). Increasing on-task behavior during mathematics independent seat-work in students with emotional disturbance by interspersing additional brief problems. *Psychology in the Schools*, 39(6), 647-659. Doi:10.1002/pits.10058
- \*Skinner, C. H., Johnson, C. W., Larkin, M. J., Lessley, D. J., & Glowacki, M. L. (1995). The influence of rate of presentation during taped-words interventions on reading performance. *Journal of Emotional and Behavioral Disorders*, 3, 214-223.
- Slavin, R. E. (1986). *Ability Grouping and Student Achievement in Elementary Schools: A Best-Evidence Synthesis*. (Rep. No.1). Baltimore, MD: Johns Hopkins University, Center for Research on Elementary and Middle Schools.
- \*Smith, B. W., & Sugai, G. (2000). Functional assessment-based behavior support planning: A self-management case study of a middle school student with EBD. *Journal of Positive Behavioral Interventions and Supports*, 2(4), 208-217.
- \*Smith, D. J., Nelson, J. R., Young, K. R., & West, R. P. (1992). The effect of a self-management procedure on the classroom and academic behavior of students with mild handicaps. *School Psychology Review*, 21, 59-72.
- Spencer, V.G. (2006). Peer tutoring and students with emotional and behavior disorders: A review of the literature. *Behavior Disorders*, 31, 204-222.
- \*Sprute, K. A., Williams, R. L., & McLaughlin, T. F. (1990). Effects of a group response cost contingency procedure on the rate of classroom interruptions with emotionally

- disturbed secondary students. *Child & Family Behavior Therapy*, 12(2), 1-12.  
Doi:10.1300/J019v12n02\_01
- \*Strong, A. C., Wehby, J. H., Falk, K. B., & Lane, K. L. (2004). The impact of a structured reading curriculum and repeated reading on the performance of junior high students with emotional and behavioral disorders. *School Psychology Review*, 33(4), 561-581.
- Suen, H. K., & Ary, D. (1989). *Analyzing quantitative behavioral observation data*. Hillsdale, NJ: Erlbaum.
- \*Sutherland, K. S., & Snyder, A. (2007). Effects of reciprocal peer tutoring and self-graphing on reading fluency and classroom behavior of middle school students with emotional or behavioral disorders. *Journal of Emotional and Behavioral Disorders*, 15(2), 103-118. Doi:10.1177/10634266070150020101
- Sutherland, K.S. & Wehby, J.H. (2001). Exploring the relation between increased opportunities to respond to academic requests and the academic and behavioral outcomes of students with EBD: A review. *Remedial and Special Education*, 22, 113-121.
- \*Swain, J. C., & McLaughlin, T. F. (1998). The effects of bonus contingencies in a classwide token program on math accuracy with middle-school students with behavioral disorders. *Behavioral Interventions*, 13(1), 11-19. Doi:10.1002/(SICI)1099-078X(199802)13:1<11::AID-BIN6>3.0.CO;2-1
- \*Swanson, H. L. (1985). Effects of cognitive-behavioral training on emotionally disturbed children's academic performance. *Cognitive Therapy and Research*, 9, 201-216.

- Swanson, H.L., & Sachse-Lee, C. (2000). A meta-analysis of single-subject design intervention research for students with learning disabilities. *Journal of Learning Disabilities, 33*(2), 114-166.
- Tankersley, M., Landrum, T. J., & Cook, B. G. (2004). How research informs practice in the field of emotional and behavioral disorders. In R. B. Rutherford, M. Quinn, S. R. Mathur (Eds.), *Handbook of research in emotional and behavioral disorders* (pp. 98-113). New York, NY US: Guilford Press.
- Tawney, J. W., & Gast, D. L. (1984). *Single Subject Research in Special Education*. Columbus, OH: Merrill.
- Templeton, J. E. (1990). Social skills training for behavior-problem adolescents: A review. *International Journal of Partial Hospitalization, 6*, 49-60.
- Templeton, T., Neel, R. S., & Blood, E. (2008). Meta-analysis of math interventions for students with emotional and behavioral disorders. *Journal of Emotional and Behavioral Disorders, 16*(4), 226-239.
- The Peacock Hill Working Group (1991). Problems and promises in special education and related services for children and youth with emotional or behavioral disorders. *Behavioral Disorders, 16*(4), 299-313.
- Theodore, L. A., Bray, M. A., & Kehle, T. J. (2004). A comparative study of group contingencies and randomized reinforcers to reduce disruptive classroom behavior. *School Psychology Quarterly, 19*, 253 – 271.
- To make education more like medicine). *Thomas B. Fordham Foundation Report of*

- Trout, A. L., Nordness, P. D., Pierce, C. D., & Epstein, M. H. (2003). Research on the academic status of children with emotional and behavioral disorders: A review of the literature from 1961 to 2000. *Journal of Emotional & Behavioral Disorders, 11*(4), 198. Doi:Article
- \*Truchlicka, M., McLaughlin, T. F., & Swain, J. C. (1998). Effects of token reinforcement and response cost on the accuracy of spelling performance with middle-school special education students with behavior disorders. *Behavioral Interventions, 13*(1), 1-10. Doi:10.1002/(SICI)1099-078X(199802)13:1<1::AID-BIN1>3.0.CO;2-Z
- \*Trussell, R. P., Lewis, T. J., & Stichter, J. P. (2008). The impact of targeted classroom interventions and function-based behavior interventions on problem behaviors of students with emotional/behavioral disorders. *Behavioral Disorders, 33*(3), 153-166.
- Trusty, J., Thompson, B., & Petrocelli, J.V. (2004). Practical guide for reporting effect size in the Journal of Counseling & Development. *Journal of Counseling and Development, 82*, 107-110.
- U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2012, March). *Children Classified as Having an Emotional Disturbance intervention report: First Step to Success*. Retrieved from <http://whatworks.ed.gov>.
- U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2012, February). *Early Childhood Education Interventions for Children with Disabilities intervention report: The Incredible Years*. Retrieved from <http://whatworks.ed.gov>.

- U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse (2011, November). *Children Classified as Having an Emotional Disturbance intervention report: The Incredible Years*. Retrieved from <http://whatworks.ed.gov>.
- U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse (2011, October). *Children Classified as Having an Emotional Disturbance intervention report: Check & Connect*. Retrieved from <http://whatworks.ed.gov>.
- U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse (2011, October). *Intervention report: Coping Power*. Retrieved from <http://whatworks.ed.gov>.
- U.S. Department of Education. (2002). *Statistical Standards*. Retrieved October 29, 2012 from <http://nces.ed.gov/statprog/2002/stdtoc.asp>
- Van den Noortgate, W., & Onghena, P. (2007). The aggregation of single-case results using hierarchical linear models. *The Behavior Analyst Today*, 8(2), 196-209.
- Vannest, K. J., Parker, R. I., Park, H., Fowler, L., Devore, H., Mohan, S., & Bullous, S. (2008). The effects of mentors on the classroom behavior of students with emotional and behavioral disorders. *Journal Of At-Risk Issues*, 14(2), 17-26.
- Wagner, M. M. (1995). Outcomes for youths with serious emotional disturbance in secondary school and early adulthood. *The Future of Children: Critical Issues for Children and Youths*, 5, 90–112
- Wagner, M., Blackorby, J., Cameto, R., Hebbeler, K., & Newman, L. (1993). *The transition experiences of young people with disabilities*. Menlo Park, CA: SRI International.



- Wagner, M., Cameto, R., & Newman, L. (2003). Youth with disabilities: A changing population: A report of findings from the National Longitudinal Transition Study (NLTS) and the National Longitudinal Transition Study-2 (NLTS2). Menlo Park, CA: SRI International.
- Wagner, M., Kutash, K., Duchnowski, A. J., Epstein, M. H., & Sumi, W. C. (2005). The children and youth we serve: A national picture of the characteristics of students with emotional disturbances receiving special education. *Journal of Emotional and Behavioral Disorders, 13*(2), 79-96.
- Wagner, M., Newman, L., Cameto, R., and Levine, P., (2005). *After High School: A First Look at the Postschool Experiences of Youth With Disabilities: A Report From the National Longitudinal Transition Study-2 (NLTS2)*. Menlo Park, CA: SRI International.
- Walker, H. M., & Severson, H. (2003). Developmental prevention of at risk outcomes for vulnerable antisocial children and youth. In K. L. Lane, F. M. Gresham, & T. O'Shaughnessy (Eds.), *Interventions for children with or at risk for emotional and behavioral disorders* (pp. 177-194). Boston: Allyn and Bacon.
- Walker, H. M., Nishioka, V. M., Zeller, R., Severson, H. H., & Feil, E. G. (2000). Causal factors and potential solutions for the persistent underidentification of students having emotional or behavioral disorders in the context of schooling. *Assessment For Effective Intervention, 26*(1), 29-39. Doi:10.1177/073724770002600105

- \*Wehby, J. H., Falk, K. B., Barton-Arwood, S., Lane, K. L., & Cooley, C. (2003). The impact of comprehensive reading instruction on the academic and social behavior of students with emotional and behavioral disorders. *Journal of Emotional and Behavioral Disorders, 11*(4), 225-238. Doi:10.1177/10634266030110040401
- Wehby, J. H., Symons, F.J., Canale, J. A., & Go, F. J. (1998). Teaching practices in classrooms for students with emotional and behavioral disorders: Discrepancies between recommendations and observations. *Behavioral Disorders, 24*, 51-56.
- Wery, J. J. & Cullinan, D. (2011). State Definitions of Emotional Disturbance. *Journal of Emotional and Behavioral Disorders.*
- White, O. R., & Haring, N. G. (1980). *Exceptional teaching* (2<sup>nd</sup> ed.). Columbus, OH: Charles Merrill.
- Whitener, E. M. (1990). Confusion of confidence intervals and credibility intervals in meta-analysis. *Journal of Applied Psychology, 75*(3), 315-321.
- Wilkinson, L., & The Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*, 594–604.
- \*Willis, R. J., Whalen, K. P., Sweeney, W. J., & McLaughlin, T. F. (1995). The effects of self-recording and contingent computer time on the off-task behavior of an adolescent with emotional disturbances. *Child & Family Behavior Therapy, 17*(1), 35-45.  
Doi:10.1300/J019v17n01\_03

Wolf, F. M. (1986). *Meta-analysis: Quantitative methods for research synthesis*. Beverly Hills, CA: Sage Publications.

**APPENDIX**

## **Appendix A**

### **Systematic Literature Reviews**

Coleman and Vaughn (2000) systematically reviewed the literature to examine what teachers' perceptions of effective reading instruction for students with emotional and behavioral disorders. The review included three group design and five single case research studies published between 1975 to 1998. Primary studies participants included elementary students (grades K through 6) with emotional and behavioral disorders. Although the term EBD was not defined, search terms included: seriously emotionally disturbed, emotional disorders, behavioral disorders). Studies of participants who were identified as solely ADHD were excluded. Results indicate teachers believe that students with EBD may not receive routine reading instruction, because of characteristics associated with EBD. However, teachers perceive them to make leaps in progress rather than slow steady progress. Teachers hypothesize that students' fear of failure and trust issues are impediments to learning. Being mindful of and reducing these burdens, and keeping students engaged is critical to reading instruction success. Cross-age tutoring is perceived to be highly motivating. Teachers of students with EBD also perceive real-world application, and explicit and direct phonics instruction as effective.

Evans, Harden, and Thomas (2004) systematically reviewed the literature to examine what strategies are effective when used with students with emotional and behavioral difficulties in mainstream primary schools. The review included 28 research studies

published between 1975 to 1999. Evans and colleagues used the English Education Ministry definition:

Emotional and behavioral difficulties range from social maladaptation to abnormal emotional stresses. They are persistent (if not necessarily permanent) and constitute learning difficulties. ... They may become apparent though withdrawn, passive, aggressive, or self-injurious tendencies (DfEE, 1994).

Students in special education classes (e.g., resource rooms) were excluded. Interventions based on behavioral, cognitive-behavioral, and systemic models were deemed to be effective for reducing off-task behavior. Interventions based on behavioral, cognitive-behavioral, and models were deemed to be effective in reducing disruptive behavior. Interventions based on cognitive-behavioral models were deemed to be effective in reducing aggressive behavior. Interventions based on cognitive-behavioral models were deemed to be effective in reducing social difficulties. Insufficient evidence was available to determine effectiveness of remaining interactions. Behavioral models were found to have positive effects for off task or disruptive behavior. Cognitive-behavioral models were found to have positive effects for off-task or disruptive behavior, and have positive immediate effects for reducing aggression or improving social skills but none showed long-term gains. Systemic (ecological) models were found to have positive effects for distractibility by changing seating location. No studies included interventions based on the psychotherapeutic model.

Gulchack and Lopes (2007) systematically reviewed the literature to examine the effect of functional assessment-based interventions for students with or at-risk for EBD as

defined by the local culture (e.g., Canada, United Kingdom). The review included 11 quasi-experimental research studies and one study reconsideration, published between 2000 and 2005. Interventions based on applied behavior analysis were generally successful in the United States, yet few other countries shared this outcome. Outside the United States holistic treatments and therapeutic nurturing environments were found to be generally effective.

Heckaman, Conroy, Fox, and Chait (2000) systematically reviewed the literature to examine the effect of functional assessment-based interventions for students with or at-risk for EBD who demonstrate problem behaviors. The review included 22 single case research studies published between 1991 and 1999. Antecedent based interventions, consequence based interventions, combination of antecedent and consequence based interventions, and all were found to be generally successful. Although the particular effects on those with EBD are not clearly delineated.

Lewis, Hudson, Richter, and Johnson (2004) conducted a systematic literature review of scientifically supported practices used with children with challenging behavior (included EBD, ADHD, at-risk and all categories of high incidence disabilities) published between 1991 and 2004. The number and type of studies not reported. Results indicated that teacher praise/reinforcement increases task engagement, decreases problem behavior. Increasing opportunities to respond (correct on-level academic responding) increases task engagement. Positive behavior support is “promising” or has emerging evidence (insufficient replications). Functional behavioral assessment-based interventions improve social behavior. Social skills

instructions are effective with students with internalizing and externalizing behaviors. Self-management improved overall behavior.

Reddy, DeThomas, Newman, and Chun (2009) reviewed the single case and group design research on school-based prevention and intervention programs by methodology employed with students with or at-risk for ED. The review included 29 primary studies published between 1988 and 2005. Deficit-based assessment and treatment approaches were investigated more than strength-based approaches. The majority of the programs used behavioral and/or cognitive behavioral orientation treatment models.

Scruggs, Mastropieri, and Richter (1985) reviewed the literature to examine the effect of peer-tutoring on students who were described as behaviorally disordered /emotionally disturbed children and youth; however included studies of students with LD, autism, educable mental retardation, or were incarcerated, or withdrawn students. The results indicated that in most cases, tutees demonstrated measureable gains in the content area being tutored. Included studies suggest mixed results of peer tutoring to improve social behavior and decrease disruptive behaviors. Control group research indicated no differences self-concept or behavior rating scales.

Spencer (2006) updated a literature review on the use of students with EBD as tutors and/or tutees to teach academic or social skills. The review included 38 primary studies published between 1972 and 2002. Findings suggest that students with EBD can be taught to serve as tutors as well as tutees.



Sutherland (2001) reviewed the literature to examine the effect of increased opportunities to respond to academic requests on the academic and behavioral outcomes of students with EBD. The review included 6 primary studies published between 1976 and 1997. The studies reviewed suggest that increased opportunities for students to respond result in higher task engagement, decreased aggressive and disruptive behavior, and academic achievement compared to fewer opportunities for students to respond. Yet, teachers rarely provide sufficient opportunities.

Zargoza, Vaughn, and McIntosh (1991) reviewed the literature to examine the effect of social skill interventions and their effects on students with behavior problems. The review included 27 primary studies published between 1980 and 1991. Participants included were school-aged students with behavior problems with at least average intelligence. Included students were those who were identified by teacher/principal as behaviorally disordered, aggressive, disruptive, isolated or unable to resolve conflict, receive education in serious emotional disturbance classroom or school, receive treatment in hospital or psychiatric setting; or a combination of settings. Results from the studies reviewed suggest that social skill interventions have been used successfully with children with behavior problems. Social skill interventions improved self, teacher and parent perceptions. Peer perceptions appeared to be more resistant to change.

## Appendix B

### Meta-analyses

Ang and Hughes (2001) conducted a meta-analysis to evaluate the differential benefits of social skill training on antisocial youth. They included 38 group and quasi-experimental design studies published between 1975 and 1999. The participants in the primary studies were school-aged (6 to 18 years old) students with antisocial behavioral problems including childhood aggression, conduct disorder, oppositional defiant disorder, antisocial behavior, violent behavior, or adolescent delinquency. They excluded studies of participants described as hyperactive, peer-rejected, or substance abusing, unless they also presented with conduct problems. The meta-analytic approach utilized Hedges'  $g$  (Hedges, 1984). The mean post-test effect was 0.62 across all included studies. Effects for heterogeneous (prosocial & deviant) student groups ( $ES=.70$ ) were higher than those with homogenous student groups ( $ES=.55$ ).

Beelman, Pflingsten, and Losel (1994) employed the meta-analytic procedure to aggregate effects of reading instruction on reading achievement with participants described as students with or at-risk of BD. While BD was not defined, search terms included: disadvantaged, behavioral disorders, emotional disturbance, problem behavior, and behavior difficulties. The meta-analysis included six experimental design and 16 single case research studies, published between 1970 and 2010, using the correlation between baseline and intervention, with Rosenthal's correction method. Students with and at-risk for BD were shown to respond to reading instruction, with  $ES$  ranging from moderate to large in both

single case research and group research. Effects were relatively consistent across instructional programs, approaches and formats.

Benner, Nelson, Ralston, and Mooney (2010) used the meta-analytic procedure to aggregate effects of reading instruction on reading achievement with participants described as students with or at-risk of BD. Participants included students with or at-risk of BD. While BD was not defined search terms included: disadvantaged, behavioral disorders, emotional disturbance, problem behavior, and behavior difficulties. The meta-analysis aggregated data from 6 group design and 18 single case research, published between 1970 and 2010, using the correlation between baseline and intervention, with Rosenthal's correction procedure. Results indicated strong effect on reading skills overall from group studies ( $g = 1.02$ ); range for SCR was ( $ES = -0.09$  to  $2.71$ ): supplemental protocol ( $ES = -0.09$  to  $2.45$ ), self-management ( $ES = 0.47$  to  $2.71$ ), and peer mediated literacy interventions ( $ES = 0.82$  to  $2.68$ ).

Durlak, Furhman, and Lampman (1991) applied the meta-analytic procedure to aggregate effects of Cognitive-Behavior Therapy with participants described as children (in groups with mean age of 13 years or younger) who have manifested some degree of behavioral or Socially Maladjusted. The meta-analysis aggregated data from 64 group research studies, published between 1970 and 1987. Results indicated stronger effect on students aged 11 to 13 ( $ES = 0.92$ ), than those aged 5 to 7 ( $ES = 0.57$ ), or 8 to 11 ( $ES = 0.55$ ). Developed cognitive processes appear critical in determining treatment responsiveness.

Gage, Lewis, and Stichter (2012) made use of the meta-analytic method to aggregate effects of functional behavioral assessment-based interventions used with students with and at-risk for EBD. The meta-analysis aggregated data from 69 single case research studies, published between 1982 and 2010. Results indicated for students with or at-risk for EBD functional behavioral assessment-based interventions reduced problem behaviors by 70.5%.

Ghafoori and Tracz (2001) conducted a meta-analysis to evaluate the success of cognitive-behavioral therapy in reducing students' disruptive behaviors. They included 27 primary studies of cognitive-behavioral therapy employing teacher measures of disruptive behavior. The primary studies, published between 1987 and 1997, compared pre-treatment to post-treatment conditions focused on decreasing disruptive behavior. The participants in the primary studies were students aged 5 to 13, that were diagnosed as having ADD/ADHD (k=3), conduct disorder (k=3), mixed diagnoses (k=1), other diagnoses (k=1) and normal or no diagnosis (k=19). The meta-analytic approach utilized Hedges'  $g$  (Hedges, 1984), and mean weighted effect size ( $d$ ) calculations (Cohen, 1988). Results showed ADD/ADHD ( $d=0.31$ ), conduct disorder ( $d=0.43$ ), mixed diagnoses ( $d=0.11$ ), and normal or no diagnosis ( $d=0.29$ ). Their findings suggest that students diagnosed with conduct disorder showed significantly greater improvements than those with mixed diagnoses.

Griffith, Hurley, and Hagan's (2009) meta-analysis examined treatment integrity of literacy interventions used with students with EBD. They included 43 primary experimental, quasi-experimental and single subject studies published between 1977 and 2005. The effect size calculations for group studies employed Cohen's  $d$  (Cohen, 1988). The

Swanson procedure (Swanson & Sachse-Lee, 2000) was used to aggregate the single subject research. Results indicated that single subject design studies were more likely to report clear operational definitions than those that used with type of group design.

Joseph and Eveleigh (2011) employed a meta-analysis to evaluate the effects of self-monitoring on reading performance of students with disabilities. This group of students included those who were identified as having learning disabilities, emotional and behavioral disorders (ED), ADHD, and speech/language disorder (although not explicitly defined). This meta-analysis included sixteen primary single subject studies and one quasi-experimental study, published between 1987 and 2008. Single subject studies were aggregated using PND (Scruggs, Mastropieri, Casto, 1987). The Cohen's  $d$  effect size calculation was used for the group study (Cohen, 1988). Students with EBD + ADHD demonstrated improved reading comprehension performance when self-monitoring was used. Students with EBD were produced PNDs with a range of 27% to 100% on accuracy measures. Effects were similar for on-task behavior (PND = 34.0) and productivity (PND = 33.0). Students with EBD performed better than those with comorbid EBD and ADHD.

Losel and Beelman (2003) conducted a meta-analysis of group design research to evaluate the effect of skills training as a universal and targeted prevention of anti-social behavior in at-risk groups. Studies of participants described as adjudicated youth and clinical groups; included conduct disorders and oppositional defiant disorders were excluded. Results include an overall estimated mean effects of  $d = 0.38$  post intervention and  $d = 0.28$  for follow-up measures. Behavioral approaches yielded a ES of  $d = 0.37$ , and a follow up ES of

$d = 0.17$ ; cognitive approaches yielded  $d = 0.39$ , and a follow up ES of  $d = 0.36$ ; cognitive behavioral approaches yielded  $d = 0.39$ , and a follow up ES of  $d = 0.37$ ; and counseling/therapy yielded  $d = 0.36$ , and a follow up ES of  $d = 0.17$ . No difference of ES for published and unpublished studies.

Mathur, Kavale, Quinn, Forness, and Rutherford (1998) conducted a quantitative synthesis (i.e., meta-analysis) to evaluate the social skill interventions used with students with emotional and behavior problems. This group of students included those with autism, ED, conduct disorder, and delinquency. Sixty-four primary single subject studies were aggregated using PND (Scruggs, Mastropieri, Casto, 1987). Social skills instruction was found to be mildly effective in increasing social interaction skills (PND=66.0), enhancing social behavior (PND=63.0), and improving communication skills (PND= 59.0).

Mooney, Ryan, Uhing, Reid, and Epstein (2005) conducted a meta-analysis to evaluate the success of self-management interventions on students with EBD. They included 22 primary single subject research reports published between 1970 and 2002 using Swanson's procedure (Swanson & Sachse-Lee, 2000). Self-instruction (ES = 2.71), multiple-component interventions (ES = 2.11), self-monitoring (ES = 1.90), strategy instruction (ES = 1.75), and self-evaluation (ES = 1.13) were all found to be effective. Interventions were also found to be effective in all academic areas: social studies (ES = 2.66), reading (ES = 2.28), math (ES = 1.97), and writing (ES = 1.13).

Nelson, Smith, Young, & Dodd (1991) executed a meta-analysis of self-management outcomes research conducted with students with behavioral disorders. Sixteen primary single

case studies published between 1976 and 2088, met the authors' inclusion criteria. The meta-analytic approach utilized PND (Scruggs, Mastropieri, Casto, 1987). Results include durable moderate to large treatment effects on social and academic behaviors. A functional relationship was found between self-instruction and increased on-task behavior (PND range 83-87%), but inconsistent generalization results leave doubt about its effectiveness. Self-evaluation with token economy was found to decrease off-task behavior effects (PND = 91%) but was not found to generalize to novel settings (PND =15%). Videotaping used to promote self-evaluation was found to be probably effective. When used in isolation self-recording PND ranged from 88 to 93%. It was shown to be less effective when it was used with reinforcement procedures moderate effect and generalized across settings (PND = 80.00).

Pierce, Reid, and Epstein (2004) conducted a meta-analysis to evaluate the success of cognitive-behavioral therapy in reducing students' disruptive behaviors. They included 27 primary studies of cognitive-behavioral therapy employing teacher measures of disruptive behavior. The primary studies, published between 1987 and 1997, compared pre-treatment to post-treatment conditions focused on decreasing disruptive behavior. The participants in the primary studies were students aged 5 to 13, that were diagnosed as having ADD/ADHD ( $k=3$ ), conduct disorder ( $k=3$ ), mixed diagnoses ( $k=1$ ), other diagnoses ( $k=1$ ) and normal or no diagnosis ( $k=19$ ). The meta-analytic approach utilized Hedges'  $g$  (Hedges, 1984), and mean weighted effect size ( $d$ ) calculations (Cohen, 1988). Results showed ADD/ADHD ( $d=0.31$ ), conduct disorder ( $d=0.43$ ), mixed diagnoses ( $d=0.11$ ), and normal or no diagnosis

( $d=0.29$ ). Their findings suggest that students diagnosed with conduct disorder showed significantly greater improvements than those with mixed diagnoses.

Quinn, Kavale, Mathur, Rutherford, and Forness (1999) conducted a meta-analysis to investigate social skill interventions used with students who meet the criteria found in the federal definition of emotional disturbance. They included 35 group design studies published between 1981 and 1994. Results indicated that social skills intervention used with individuals or small groups is not very effective in increasing social competence of students with EBD. Often the type of social skill deficit was not considered in the intervention selection. Formative evaluations should guide the selection and design of social skill interventions. Too often, interventions were not implemented long enough to be effective.

Ryan, Reid, and Epstein (2004) examined the effectiveness of teacher-mediated interventions on the academic functioning of students with EBD using Swanson's procedure (Swanson & Sachse-Lee, 2000) aggregating single subject research and Cohen's  $d$  to aggregate the group design research. Thirty primary studies of antecedent- or consequence-focused teacher-mediated interventions, published between 1968 and 2001. Results indicated positive academic outcomes ( $d = 1.05$ ) across all teacher-mediated interventions. Antecedent-focused interventions had a higher average mean effect size ( $d = 1.31$ ) than did consequent-focused interventions ( $d = 0.80$ ).

Schneider and Byrne (1985) conducted a meta-analysis to investigate social skill interventions used with students identified as "normal", sociometrically assessed, withdrawn students identified as ED, aggressive identified as ED, learning disabled, and mentally



retarded. Students with other ED characteristics were excluded. Fifty-one group design studies published between 1969 and 1983, met the authors' inclusion criteria. Results indicated higher effects for operant procedures ( $ES = 0.85$ ) and modeling ( $ES = 0.75$ ) than for coaching ( $ES = 0.65$ ) and social cognitive approaches ( $ES = 0.55$ ).

Spencer (2006) conducted a meta-analysis to investigate the effects of peer tutoring used with students identified as having emotional or behavioral disorders, which included students identified with emotional behavioral disorders (about 54%), serious emotional disturbance, or as being juvenile delinquents or emotionally handicapped, or as exhibiting maladaptive behavior. Fifteen group design, seven quasi-experimental, and 14 single case studies published between 1972 and 2003, met the authors' inclusion criteria. Results indicated higher effects cross-age tutoring in reading ( $d = 0.82$ ), than in spelling ( $d = 0.37$ ) and math ( $d = 0.21$ ). Larger effects for same-age tutoring on math ( $d = 0.36$ ) than in reading ( $d = 0.21$ ). Larger effects for reciprocal tutoring on social studies ( $d = 0.84$ ) than in reading ( $d = 0.73$ ).

Templeton, Neel, and Blood (2008) conducted a meta-analysis to evaluate math interventions used with students with EBD. Fifteen primary single case studies published between 1984 and 2002, met the authors' inclusion criteria. The meta-analytic approach utilized PND (Scruggs, Mastropieri, Casto, 1987). Approximately half of the included studies reported maintenance and generalization measures, and were successful and very effective on average ( $PND = 97.78$ ). Interventions with math as a primary interest ( $PND = 87.30$ ) were more significantly effective than those interventions with math as a secondary

focus (e.g., time on task; PND = 62.46). Interventions aimed to improve math fact recall (PND=89.70) although not significant, were higher than those aimed at improving calculation (PND=80.89). Interventions including strategy instruction (PND=86.41) were not significantly different than those that did not include strategy instruction (PND = 80.50). Interventions without environmental accommodations had higher PNDs (PND = 91.10) than those with accommodations (PND = 75.32)

## Appendix C

### Studies Excluded

#### *Study Did Not Investigate An Intervention*

---

Bauer, Sterba, & Hallfors (2008)  
 Fore, Boon, Burke, & Martin (2009)  
 Gary (2010)  
 Gratz, Rosenthal, Tull, Lejuez, & Gunderson (2009)  
 Gunter, Venn, Patrick, Miller, & Kelly (2003)  
 Heckaman, Conroy, Fox, & Chait (2000)  
 Hodges, Xue, & Wotring (2004)  
 Levick (1988)  
 Luth (2001)  
 McKie & Mathai (1987)  
 Megone (2010)  
 Miller, Heafner, & Massey (2009)  
 Miller, Williams, & McCoy (2004)  
 Muir & Elisabeth (1985)  
 Reitz (1994)  
 Rizza & Morrison (2003)  
 Segal & Swallow (1993)  
 Segal, King, & Naylor (1995)  
 Shapiro & Sandra (2008)  
 Storch, Lehmkuhl, Ricketts, Geffken, Marien, & Murphy (2010)  
 Tissue & Korz (1993)  
 Wassef, Collins, Ingham, & Mason (1995)  
 Wells & Matthews (1996)  
 Wiedermann & Wiedermann (1988)  
 Wiley, Siperstein, Bountress, Forness, & Brigham (2008)  
 Williams & Hollis (1999)  
 Williams & Hollis (1999)  
 Zimet & Farley (1987)  
 Zimet & Farley (1986)

*Study Participants Did Not Meet The Definition Of ED*

---

- Anderson-Butcher, Newsome, & Nay (2003)  
August, Lee, Bloomquist, Realmuto, & Hektner (2004)  
Banks & Zionts (2009)  
Bardon, Dona, & Symons (2008).  
Belfiore, Lee, Vargas, & Skinner (1997)  
Caselman (2005)  
Conley, Caldarella, Young (2007)  
Diken & Rutherford (2005)  
Ducharme, Folino, & DeRosie (2008)  
Harris, Oakes, Lane, & Rutherford (2009)  
Ingram, K., Lewis-Palmer, T., & Sugai, G. (2005)  
Kodak, Fisher, Clements, & Bouxsein (2011)  
Mace & Heller (1990)  
McIntosh & Dodd (2008)  
McKissick, Hawkins, Lentz, Hailley, & McGuire  
Miller, Heafner, & Massey (2009)  
Mozzoni & Hartnedy (2000)  
Nahgahgwon, Umbreit, Liaupsin, Turton (2010)  
Nelson, Smith, & Colvin (1995)  
Nelson, Stage, Epstein, & Pierce (2005)  
Newcomer & Lewis, (2004)  
Oakes, Mathur, & Lane (2010)  
Park & Scott (2009)  
Put, Van den Bergh, Demedts, & Verleden (2000)  
Simons, Mynors-Wallis, Pickering, Gray, Brooking, & Thompson (2001)  
Stage, Jackson, Moscovitz, Erickson, Thurman, Jessee, & Olson (2006)  
Wilczynski, Fusilier, Dubard, & Elliott (2005)  
Williams, Hermens, Palmer, Kohn, Clarke, Keage, Clark, & Gordon (2008)  
Williams, Kirkpatrick-Sanchez, Enzinna, Dunn, Borden-Karasack (2009)

*Study Did Not Occur In A School Environment*

---

- Adelman, Kaser-Boyd, & Taylor (1984)  
Appelstein (1993)  
Apsche, Bass, & Siv (2006)  
Baker-Ericzen, Hurlburt, Brookman-Fraze, Jenkins, & Hough (2010)  
Dale, Baker, Anastasio, & Purcell (2007)  
Gleser & Lison (1986)  
Goulding (1991)  
Graves, Shelton, & Kaslow, (2009)  
Heflinger & Taylor-Richardson (2004)  
Herbert, Gaudiano, Rheingold Myers, Dalrymple, & Nolan (2005)  
Hoffart, Versland, & Sexton (2002)  
Ilyukhina, Bokarius, Matveev, & Rummyantseva (1994)  
Jones (1992)  
Morgan (2010)  
Rolvsjord, (2001)  
Scarlato & Asahara (2004)  
Shirk, Kaplinski, & Gudmundsen (2009)  
Smith, Handler, & Nash (2010)  
Wilhelm (2004)  
Wilholm (1984)  
Williams (1993)  
Williams & Evans (2003)  
Williams & Sherr (2009)  
Williams & Spruill (2005)  
Yeo, Wong, Gerken, & Ansley (2005)  
Zarski & Fluharty (1992)  
Zeiger (1994)

*Study Did Not Investigate Student Outcomes*

---

Allen (2000)  
Alter, Conroy, Mancil, & Haydon (2008)  
Badura & Steinmeyer (1984)  
Chiodo (1987)  
Chosak, Marques, Fama, Renaud, & Wilhelm (2009)  
Clark & Elliott (1988)  
Coddling, Feinberg, Dunn, & Pace, (2005)  
Davis & Lysaker (2005)  
Farran, Gilley, McCann, Bienias, Lindeman, & Evans (2007)  
Fogt, George, Kern, White, & George (2008)  
Kalis, Vannest, & Parker (2007)  
O'Connor, Koszegi, Aardema, van Niekerk, & Taillon (2009)  
Papageorgiou & Wells (2000)  
Penn & Combs (2000)  
Reinecke & Hoyer (2010)  
Singh, Lancioni, Winton, Singh, Adkins, & Singh (2009)  
Snell, Voorhees, & Chen (2005)  
Upton, Bundy, & Speed (1986)  
Wiklund, Sanne, Vedin, & Wilhelmsson (1984)  
Williams, Roth, Kuzniecky, & Faught (1994)

*Studies that did not take place in the US or Canada*

---

Baker, Archer, & Curtis (2005)

Fabre, (1992)

Höger, Temme, Reiter, & Steiner (1994)

Mowat, (2010)

Schneider (1993)

Schuppert, Giesen-Bloo, van Gemert, Wiersema, Minderaa, Emmelkamp, & Nauta (2009)

Warren, McLellam, & Ponzoha, (1988)

Whicher, (2005)

Zonneville-Bender, Matthys, van de Wiel, & Lochman (2007)

## Appendix D

### Tables

Table 1

*Codification of Independent and Dependent Variables*

Participants and Setting	
ED label	Label used to describe the emotional disturbance of the participants was coded; additional diagnoses
Type emotional or behavioral problem	<p>Table 4 Inability to Learn:</p> <ol style="list-style-type: none"> <li>1. poor academic performance (e.g., math, reading, spelling);</li> <li>2. weak listening or note-taking skills;</li> <li>3. inadequate ability to complete independent work;</li> <li>4. easily distracted (e.g., does not pay attention to teacher or work);</li> <li>5. lacks interest, motivation, or positive attitude toward school tasks; or (f) dependency on adults or peers to complete school tasks.</li> </ol> <p>(B) Relationship problem:</p> <ol style="list-style-type: none"> <li>1. has few friends;</li> <li>2. does not work well in a group;</li> <li>3. rejected or avoided by peers;</li> <li>4. picked on or persecuted by other students;</li> <li>5. lacks social skills; or</li> <li>6. avoids interacting with people.</li> </ol> <p>I Inappropriate Behavior:</p> <ol style="list-style-type: none"> <li>1. disrespectful or defiant of authority;</li> <li>2. cruel to peers;</li> <li>3. verbally aggressive, teases or taunts others;</li> <li>4. fails to consider the consequences of own acts;</li> <li>5. cheats, lies or steals;</li> <li>6. makes threats;</li> <li>7. is loud, disruptive or rowdy;</li> <li>8. destroys or ruins things;</li> <li>9. physically assaults or fights; or</li> <li>10. uses obscene, profane, or sexually oriented language</li> <li>11. diagnosed with oppositional defiant disorder or conduct</li> </ol>



disorder

(D) Unhappiness or Depression:

1. lacks self-confidence;
2. has feelings of worthlessness;
3. experiences little joy or pleasure;
4. sad much of the time;
5. rarely smiles;
6. little or no interaction with teachers;
7. loss of interest in formerly enjoyed experiences;
8. pessimistic about the future;
9. makes statements that indicate sadness, dejection, despair or gloom; or
10. appears to lack self-confidence.

(E) Physical Symptoms or Fears:

1. complains of physical discomfort;
2. anxious, worried, or tense;
3. afraid of unlikely dangers or calamities;
4. talks of suicide or own death;
5. overly sensitive feelings and emotions;
6. feels excessively guilty;
7. shows nervous habits (e.g., tics, bites nails, twists hair); or
8. harms own body (e.g., picks self, cuts self, writes on self).

Socially Maladjusted:

1. vandalizes property;
2. participates in illegal, anti-social, or gang activities;
3. abuses drugs or alcohol before or after school;
4. exhibits precocious sexual behavior;
5. runs away from home; or
6. steals.

Schizophrenia-Like Features:

1. Schizophrenic or having Schizophrenia;
2. thought problems; or
3. psychotic.

Age and grade

When grade is not provided, those in Kindergarten will be assumed to be 5 years old, first grade 6 years old, and so forth.

If only grade level was reported, the same conversion was conducted in reverse.

Sex

- (a) male,
- (b) female, or
- (c) not determinable.

Educational

- (a) public,

placement	(b) private school, or (c) unreported.
Special education placement	(a) regular is defined as receiving services less than 21% out of regular class, (b) resource 21%—60% out of regular class, I Self-Contained greater than 60% out of regular class but in regular school (as well as settings described as “separate program” unless it is described as located in a separate building or facility) (d) alternative greater than 50% of day in alternative or day-treatment school, (e) residential or hospital setting, or (f) unreported.

---

#### Study Characteristics

Single case research design	(a) withdrawal, (b) alternating treatment, simultaneous treatments, or (d) multiple baseline (MB) designs: (d <sub>1</sub> ) across settings, (d <sub>2</sub> ) across participants, (d <sub>3</sub> ) across probes, or (d <sub>4</sub> ) across behaviors.
Quality of evidence: IOA in each phase	(a) recorded inter-observer agreement on at least 20% of observations in each phase (b) did not meet this criteria
Quality of evidence: IOA minimum thresholds	(a) inter-observer agreement met minimum thresholds (see Suen & Ary, 1989 for thresholds) (b) did not meet this criteria
Quality of evidence: phase repetitions	Withdrawal design (a) had at least four different phases (b) did not meet this criteria Multiple baseline (a) design had at least six different phases (b) did not meet this criteria Alternating and simultaneous treatment (a) designs had at least five data points per phase, (b) did not meet this criteria

Quality of evidence: data points per phase	<p>Withdrawal design</p> <ul style="list-style-type: none"> <li>(a) with at least five data points per phase</li> <li>(b) at least four phases with at least three data points per phase</li> <li>(c) did not meet this criteria</li> </ul> <p>Multiple baseline design</p> <ul style="list-style-type: none"> <li>(a) with at least five data points per phase.</li> <li>(b) with at least three data points per phase,</li> <li>(c) did not meet this criteria</li> </ul> <p>Alternating and simultaneous treatment designs</p> <ul style="list-style-type: none"> <li>(a) with least five data points per phase,</li> <li>(b) at least three data points</li> <li>(c) did not meet this criteria</li> </ul>
--	--

---

#### Outcome Variables

---

Academic or behavioral	<p>Academic outcomes included both achievement data, learning behavior outcomes (e.g., on-task behavior).</p> <p>Behavioral outcomes measured observable behaviors (e.g., hitting, swearing).</p>
------------------------	---

Interventions	<ul style="list-style-type: none"> <li>(a) choice (i.e., students had a choice of reinforcement);</li> <li>(b) use of procedures (i.e., routines);</li> <li>(c) functional based interventions (i.e., interventions developed to meet the need or function of the undersirable behavior);</li> <li>(d) positive reinforcement (i.e., the application of a pleasant stimulus in response to a desired behavior including praise, desirable activity, and edibles);</li> <li>(e) negative reinforcement (i.e., the removal of an unpleasant stimulus in response to a desired behavior);</li> <li>(f) punishment (i.e., the application of an unpleasant stimulus in response to an undesired behavior);</li> <li>(g) self-management/monitoring (e.g., video feedback, self-reinforcement, self instruction, self-talk, mindfulness);</li> <li>(h) peer modeling (i.e., live peer demonstrating a desirable behavior); video modeling (i.e., recorded demonstrating a desirable behavior, by the participant or another person); and</li> <li>(i) social skill instruction (i.e., instruction aimed at improving social skills)</li> <li>(j) instruction (teaching or education);</li> <li>(k) computer-assisted instruction (teaching or education with a computer);</li> <li>(l) direct instruction (explicit teaching of a skill);</li> <li>(m) opportunities to respond (the provision of multiple</li> </ul>
---------------	--

---

- occasions for student response);
- (n) curricular modifications (adjustment of educational activities/materials);
  - (o) learning strategy instruction (explicit teaching of a process to improve learning); and
  - (p) peer tutoring (education by a student).

Note: Token economies with response cost systems were coded as a combination of positive reinforcement and punishment

Data Collection	
PAND calculation	(a) number of data points in baseline (b) treatment phases, (c) minimum data points that non-overlap

Table 2

*Included SCR Studies*

Study	Intervention type (s)	Dependent Variable	Outcomes and effects
Alber, Anderson, Martin, & Moore (2005)	Instruction, Positive Reinforcement, Self-monitoring	Appropriate recruiting teacher attention, inappropriate recruiting, completion of math assignments,	Students with EBD can be taught to recruit positive attention. A functional relationship exists between decreased inappropriate recruitment and appropriate recruitment.
Anthony, Wolery, Werts, Caldwell, & Snyder, (1996)	Instruction with and without daily instructive feedback stimuli	Correct naming states and landmarks, when presented states' capital city	Students' percent of correct responses increased during instruction. These gains were maintained.
Barton-Arwood, Wehby, & Falk (2005)	Direct Instruction ( <i>Horizons Fast Track</i> reading program), Peer-Assisted Learning Strategies (PALS)	Phoneme blending, Phoneme segmentation, Nonsense word fluency, Word reading, Oral reading fluency, Total Inappropriate Behavior, Engagement	Phoneme blending increased, Phoneme segmentation increased, Nonsense word fluency were less evident rates of improvement, Word reading had positive but variable trends, Oral reading fluency positive but variable trends, Total Inappropriate Behavior no clear change between phases, Engagement no clear change between phases
Beare, Torgerson, & Creviston (2008)	Positive reinforcement	Verbal response, words spoken per minute	Positive reinforcement was effecting in increasing verbal response, words spoken per minute
Bell, Young, Blair,	Class-wide peer	History test question	Immediate and substantial gains with decreased

& Nelson (1990)	tutoring	responses	variability of test scores.
Billingsley, Scheuermann & Webber (2009)	Direct Instruction, Computer assisted instruction	Curriculum based math assessments	Neither condition was clearly superior for all participants, but both interventions were shown to have a very high effect on student performance
Blankenship, Ayres, & Langone (2005)	Computer assisted instruction	Percent correct on section and chapter quizzes	Upward trend indicates acquisition increased over time.
Bowman-Perrott, Greenwood, & Tipia (2007) Experiment 2	Class-wide Peer Tutoring	Mean percent correct On-task behavior (observation)	Class mean increased over baseline.
Brasch, Williams, & McLaughlin (2008)	Direct instruction	Number of correct multiplication facts	A positive functional relationship between mastery and retention of multiplication facts with intervention.
Callicott & Park (2003)	Self-management, Positive Reinforcement	Student speed and accuracy of math problem-solving, attending behavior/task engagement	Some evidence of effect of self-talk on academic performance.
Cancio, West, & Young (2004) Carr & Punzo (1993)	Self-management Self-monitoring	Homework completion, homework accuracy Academic accuracy and productivity	strong evidence of immediate improvement in homework completion and accuracy. Self-monitoring of academic accuracy and productivity during reading, spelling and math

Cieslar, McLaughlin, & Derby (2008)	Copy, Cover, and Compare	Math and spelling accuracy	improved academic performance. A clear functional relationship between Copy, Cover, and Compare and math and spelling performance.
Daly, Garacz, Olson, Persampieri & Ni (2006)	Instruction, Positive reinforcement	Oral reading fluency	Instructional antecedents with reinforcement contingencies increased oral reading fluency rates.
Ducharme & Harris (2005)	Error-less embedding with rapport-based support,	On-task behavior, off-task behavior	A functional relationship exists between error-less embedding and increased independent work behavior.
Dunlap, White, Vera, Wilson & Panachek (1996)	Curriculum modifications	Task engagement; Problem behavior	Functional assessment and curricular revision are a proactive intervention for reducing problem behaviors.
DuPaul & McGoey (1997)	Self-management, token reinforcement	Positive and negative behavior	Found an inconsistent relationship between self-management and token reinforcement, and behavior change.
Faulk & Wehby (2001)	Instruction, Peer Tutoring	Letter-sound association, segmentation, and blending	Peer-tutoring was effecting in increasing student performance on letter-sound identification and blending.
Ferguson, Ashbaugh, O'Reilly, & McLaughlin (2008)	Positive reinforcement	Length of transition time	Training students to respond to teacher prompts with reinforcement for appropriate responding proved to be effective to decrease the amount of time spent in transition from one activity to another.
Foxx, Kyle, Faw,	Social skill instruction	Stating three criterion	Trained students were able to maintain and

& Bittle (1989)	with problem solving strategy	questions	generalize problem-solving skills.
Freeman & Dexter-Mazza (2004)	Self-monitoring; self-monitoring + Reinforcement	Off-task and disruptive behavior	Self-monitoring with teacher matching produced improvements in disruptive and off-task classroom behavior.
Gulchak (2008)	Self-monitoring with hand-held computer	On task behavior	A functional relationship exists between a student's use of a hand-held computer to self-monitor attention and improved on-task behavior.
Gunter, Miller & Venn (2003)	Self-monitoring	Words read/minute	Self-graphing of academic performance on desk-top computer was shown to improve reading fluency.
Hale, Skinner, Winn, Oliver, Allin, & Molloy (2005)	Listening, listening while reading (Curriculum modifications)	Reading comprehensions	Listening while reading improved comprehension rates across all participants.
Hansen & Lignugaris-Kraft (2005)	Social skills instruction,	Positive verbal interactions; negative verbal interactions	Found social skill training on giving positive statements to peers, paired with dependent group contingency increased and maintained positive statements emitted to peers and decrease the rate of negative statements.
Hawkins & Axelrod (2008)	Function Based Assessment,	On-task behavior	Results supported the hypothesis that off-task behavior during homework was maintained by



	Reinforcement		escape from the task. Differential reinforcement increased on-task behavior.
Hughes, Deshler, Ruhl, & Schumaker (1993)	Direct Instruction	Improved test taking/test scoring	Students were able to acquire and maintain use of a comprehensive test-taking strategy. Students mastered four guessing strategies and apply complex series of test-taking strategies.
Jolivette, Wehby, Canale, & Massey, (2001)	Choice making	Student engagement; Disruption; Off-task behavior, math performance	Choice-making opportunities effectively improved task-engagement and reduced off-task and disruptive behaviors, and increased math performance and accuracy.
Kern, Childs, Dunlap, Clarke, & Falk (1994)	Curricular modifications, Functional-based assessment	On-task behavior	Curricular modifications were shown to be an effective intervention.
Kern, Delaney, Clarke, Dunlap, Childs (2001)	Choice of medium	Task engagement	Found functional assessment and provide information about simple classroom adaption for student that can be influential in increasing task engagement and decrease challenging behavior.
Kern, Dunlap, Childs, & Clarke, (1994)	Classwide Self-management	On-task Disruptive behavior	Demonstrated the effectiveness of self-monitoring procedures for improving on-task behavior of students.
Kern, Wacker, Mace, & Falk, (1995)	Self-evaluation (video feedback) + rewards; self-evaluation only	Inappropriate and appropriate peer interactions	Self-evaluation with video-tape procedures can improved behavior, videotaped feedback with rewards are more effective the rewards alone to

Experiment 1			improve peer-to-peer interactions
Kern, Wacker, Mace, & Falk, (1995)	Self-evaluation (video feedback) + rewards; rewards only; self-evaluation only	Inappropriate peer interactions; appropriate peer interactions	Self-evaluation with rewards are more effective to decrease inappropriate peer-to-peer interactions.
Experiment 2 Kern-Dunlap, Dunlap, Clarke, Childs, White, & Stewart (1992)	Video feedback	Desirable behavior; Undesirable behavior	This study demonstrated the efficacy of video feedback package for increasing the ratio of desirable behaviors to undesirable behaviors.
Lane, Eisner, Kretzer, Bruhn, Crnabori, Funke, Lerner, & Casey (2009)	Functional Behavioral Assessment	Academic engagement	A functional relationship was found between functional behavioral assessment and increased academic engagement.
Lawson & Greer (2006). Experiment 1	Writer Immersion and self-editing (Instruction)	Number and accuracy of structural components of writing,	Reader/writer intervention was found to increase the number of sentences written and the percent of accurate structural components included, but did not increase number of adjectives and adverbs used, the number of novel sentence frames, nor the number of components drawn by the reader.
Lawson & Greer (2006). Experiment 2	Instruction	Structural components of writing, Functional effects of writing	Writer immersion intervention was found to increase the functional and structural components of writing, and improve the accurate use of structural components.
Lloyd & Hilliard	Self-recording (Self-	On-task behavior	Students may respond differently to self-

(1989)	monitoring),		reinforcement contingency after repeated exposure to that contingency. Students who recorded their behavior accurately during the self-reinforcement condition failed to do so in the later phases of that condition.
Locke & Fuchs (1995)	Peer tutoring	On task behavior; Social behavior	Results indicated peer tutoring exerted a controlling effect on on-task behavior and social interaction.
Maher (1986)	Peer tutoring	Homework completion; Homework accuracy; attendance	Cross-age tutoring was found to increase tutee completion of academic work, and mean items correct; and increase tutors' attendance rates and reduce disciplinary referrals.
McCurdy, Cundari, & Lenz (1990)	Time delay, trial and error, Peer modeling, instruction	Words read correctly,	Time-delay may be more effective than trial and error. Both peer modeling, and instruction were effective, with little difference between them.
McCurdy & Shapiro (1988)	observing peer, self-observation	Words read correctly (2 syllable words)	Idiosyncratic treatment effects across participants.
McKenny & Dattilo (2001)	Social Skill Instruction	Encouraging; helping; conflict resolving; physical aggression; verbal aggression	Social skill instruction may have had a limited effect on prosocial behavior, but did not influence their anti-social behavior. Encouraging and helping behaviors appeared to improve as a result of the intervention; however these behaviors were not maintained.

McLaughlin (1992)	Positive Reinforcement	Percent correct in reading	Reading accuracy improved when written feedback was provided.
Miller & Cole (1998)	Social skills intervention	Giving compliments; Offering help to others; Responding to teasing	Social skill intervention resulted in increased demonstration of prosocial skills.
Moore, Cartledge, & Heckaman, (1995)	Social skills instruction; Self-monitoring	Inappropriate peer-reactive behavior; Appropriate peer-reactive behavior; Inappropriate reactions to losing; Appropriate reactions to losing; Inappropriate reactions to winning; Appropriate reactions to winning.	Results showed direct social skill instruction and self-monitoring of appropriate and inappropriate reactions to peers had a positive effect on target behaviors.
Mulcahy & Krezmien (2009)	Instruction	Math probes	Contextualized instructional geometry and measurement package produced positive effects on math accuracy.
Musser, Bray, Kehle, & Johnson (2001)	Problem Analysis Interview (Positive reinforcement, Punishment)	Disruptive behavior	Multi-component intervention that incorporates precision requests, mystery motivators, token economy with response cost, and antecedent strategies may be useful to reduce disruptive behaviors.
Nelson & Behler (1989)	Self-instruction, Peer tutoring	On task; Off task;	

		“other” maybe errors Teacher rating scale Self-control rating scale	
Ninness, Ellis, Miller, & Baker (1995)	Self-assessment	Aggressive behavior	Students can be taught to use self-management techniques to control previously performed maladaptive behaviors without adult supervision.
Ninness & Fuerst (1995)	Self-management	Off-task/disruptive behavior	Found no consistent difference in performance between baseline and self-management intervention, although there was a slight and temporary decline in between-class problem behavior.
Ogier & Hornby (1996)	Positive reinforcement	Out of seat; Talking out of turn; Not on task	Disruptive behaviors were substantially reduced through differential reinforcement. There was some evidence for the maintenance of these behaviors.
Patterson (2005)	Guided notes	Academic performance: Note recording; Quiz score	Guided notes can be an effective intervention for improving academic performance and positive outcomes.
Popkin & Skinner (2003)	Positive Reinforcement	Academic performance (spelling, math, English)	Positive reinforcement produced immediate, stable and educationally valid increases in spelling and math performance.
Prater, Hogan, & Miller, (1992)	Self-monitoring	On task behavior; Academic scores	Students can be taught to self-monitor and apply it to mainstream settings and increase on-task performance, academic accuracy.

Rafferty & Raimondi (2009)	Self-monitoring	On-task behavior; Academic performance – Math (addition and subtraction problems)	Self-monitoring of performance may be more effective than self-monitoring of attention in producing higher levels of social and academic behavior during math practice.
Rasmussen & O’Neill (2006)	Positive reinforcement	Verbal disruptions	Positive reinforcement produced variable but lower rates of verbal disruption.
Salend, Reynolds, & Coyle, (1989)	Good Behavior Game (Positive Reinforcement)	Inappropriate Behavior (verbalizations; sweating/negative comments); Drumming/touching	The Good Behavior Game is an effective strategy for simultaneously decreasing Inappropriate Behaviors.
Salend & Sonnenschein (1989)	Learning Together Approach (cooperative learning –peer modeling)	On-task behavior; cooperative behavior; academic performance	Learning Together Approach is an effective method for increasing on-task behavior, group cooperation, and academic performance.
Schoenfeld & Mathur 2009	FRIENDS (Positive reinforcement)	anxiety, behavioral performance, and academic engagement	Cognitive Behavioral Interventions was shown to have positive effects on anxiety, behavioral performance, and academic engagement.
Singh, Lancioni, Singh Joy, Winton, Sabaawi, Wahler, et al. (2007)	Mindfulness Training (Self-monitoring)	Aggression, fire-setting, cruelty, noncompliance or bullying	Students can self-regulate their behaviors in setting of their choice for as long as they wish.
Skinner & Belfiore, (1992)	Cover, Copy Compare (Instruction)	States labeled correctly with postal abbreviation	Cover, Copy Compare could improve academic performance.

Skinner, Hurst, Teeple, & Meadows (2002)	Instruction	On-task behavior during math seatwork	Interspersing brief problems increased problem completion rates.
Skinner & Johnson (1995)	Instruction	number of words read correctly, reading fluency,	Taped word interventions may be effective in increasing word list reading performance
Smith (B), Sugai, & Brown (2000)	Self-management	On-task; Talk outs; Other off task	Self-management with functional behavioral assessment was effective in increasing on-task behavior, and reducing other off-task behaviors.
Smith (D) & Young (1992)	Self-management	Academic performance Off task	Self-management was shown to reduce disruptive behaviors, and positively impact both quality and quantity of student work; however these effects did not generalize across settings.
Sprute, Williams, & McLaughlin (1990)	Response Cost (Positive reinforcement, Punishment)	Classroom interruptions	Response cost procedures substantially decreased the frequency of classroom interruptions
Strong, Wehby, Falk, & Lane (2004)	<i>Corrective Reading</i> (Direct Instruction)	Words read per minute; comprehension	Students experienced moderate growth in oral reading fluency during the <i>Corrective Reading</i> Program. Inconsistent comprehension improvement.
Sutherland & Snyder (2007)	Peer tutoring	Disruption per session; Active responding; Reading fluency	Reciprocal peer tutoring and graphing was shown to decrease disruptive behavior, increase active responding.
Swain & McLaughlin	Token Reinforcement (Positive	Math problems	Token reinforcement had a positive effect on academic responding.

(1998)	reinforcement)		
Swanson (1985)	Individual and group contingency, with response cost (Positive reinforcement, Punishment)	Positive social interaction; Negative social interaction	Group contingency was effective in improve the social interaction of classroom dyads for same- and opposite-sex interactions. Group contingencies were more effective in the promotion of social interaction than individual contingencies.
Theodore, Bray, Kehle, & Jenson (2001)	Group contingency and reinforcement	Disruptive behaviors	Group contingency with reinforcement had an immediate and dramatic effect on decreasing disruptive classroom behavior.
Trulicka, McLaughlin, & Swain (1998)	Token Economy (Positive reinforcement, Punishment)	Spelling	Token reinforcement with response cost inconsistently increased the accuracy of performance in spelling.
Trussell, Lewis, & Stichter (2008)	Function behavior intervention	Problem behaviors	Functional Behavioral Assessment (with targeted classroom changes) were sufficient in reducing problem behavioral across all participants.
Wehby, Falk, Barton-Arwood, Lane, & Cooley (2003)	Instruction ( <i>Open Court Reading</i> ) Peer Tutoring	Academic and social behavior: Correct nonsense words; Correct blends; Correct letter sounds; Correct sight words Time attending;	<i>Open Court Reading</i> and Peer Tutoring were moderately effective in improving some of the students performance measures of sound naming, blending and nonsense words. Gains in phonological awareness did not generalize to reading ability and little improvement in sight word reading was noted.



		Inappropriate Behavior	
Wilkinson (2005)	Conjoint Behavior Consultation (Function behavior assessment) and self-management	On-task and compliant behavior	Conjoint Behavior Consultation and self-management was associated with immediate and distinguishable improvements in ratings of behavioral control (on-task and compliant behavior) and effects were maintained over time.
Willis, Whalen, Sweeney & McLaughlin (1995)	Self-management	On task behavior; Off task behavior	Self-recording and contingent computer time was effective in decreasing the frequency of off-task behaviors.

---

Table 3

*Individual Study and Comparison Effective Sizes*

Author(s), Year	PAND	RPhi	Confidence Interval
<b>Alber, Anderson, Martin, &amp; Moore, 2005</b>	<b>97.7%</b>	<b>0.952</b>	<b>[0.899, 0.980]</b>
1	100.0%	1.000	[0.242, 1.000]
2	90.9%	0.771	[-0.037, 1.000]
3	95.2%	0.869	[0.277, 1.000]
4	100.0%	1.000	[0.387, 1.000]
5	100.0%	1.000	[0.480, 1.000]
6	96.2%	0.915	[0.477, 1.000]
7	100.0%	1.000	[0.451, 1.000]
8	77.8%	0.550	[-0.024, 0.882]
9	95.2%	0.869	[0.277, 1.000]
10	100.0%	1.000	[0.242, 1.000]
11	100.0%	1.000	[0.153, 1.000]
12	100.0%	1.000	[0.424, 1.000]
13	100.0%	1.000	[0.387, 1.000]
14	100.0%	1.000	[0.480, 1.000]
15	100.0%	1.000	[0.605, 1.000]
16	100.0%	1.000	[0.451, 1.000]
17	100.0%	1.000	[0.362, 1.000]
18	100.0%	1.000	[0.424, 1.000]
19	100.0%	1.000	[0.042, 1.000]
20	100.0%	1.000	[0.257, 1.000]
<b>Anthony, Wolery, Werts, Caldwell, &amp; Snyder, 1996</b>	<b>97.0%</b>	<b>0.667</b>	<b>[0.133, 0.938]</b>
1	100.0%	1.000	[0.493, 1.000]
2	95.8%	0.914	[0.471, 1.000]
3	96.0%	0.920	[0.500, 1.000]
4	96.3%	0.921	[0.505, 1.000]
<b>Barton-Arwood, Wehby, &amp; Falk, 2005</b>	<b>91.8%</b>	<b>0.834</b>	<b>[0.777, 0.878]</b>
1	90.5%	0.738	[0.149, 0.977]
2	94.4%	0.862	[0.241, 1.000]

3	72.0%	0.417	[-0.056, 0.760]
4	76.9%	0.512	[0.045, 0.820]
5	95.2%	0.869	[0.277, 1.000]
6	83.3%	0.308	[-0.204, 0.773]
7	88.0%	0.750	[0.286, 0.954]
8	84.6%	0.675	[0.210, 0.915]
9	100.0%	1.000	[0.424, 1.000]
10	100.0%	1.000	[0.397, 1.000]
11	80.0%	0.583	[0.107, 0.866]
12	84.6%	0.675	[0.210, 0.915]
13	91.9%	0.761	[0.327, 0.956]
14	97.0%	0.917	[0.494, 1.000]
15	100.0%	1.000	[0.949, 1.000]
16	95.1%	0.895	[0.585, 0.991]
17	97.3%	0.920	[0.510, 1.000]
18	100.0%	1.000	[0.586, 1.000]
19	100.0%	1.000	[0.596, 1.000]
20	96.2%	0.917	[0.484, 1.000]

**Beare, Torgerson, &  
Creviston, 2008**

	<b>100.0%</b>	<b>1.00</b>	<b>[0.957, 1.000]</b>
1	100.0%	1.000	[0.192, 1.000]
2	100.0%	1.000	[0.407, 1.000]
3	100.0%	1.000	[0.139, 1.000]
4	100.0%	1.000	[0.468, 1.000]
5	100.0%	1.000	[0.665, 1.000]
6	100.0%	1.000	[0.486, 1.000]
7	100.0%	1.000	[0.192, 1.000]
8	100.0%	1.000	[0.407, 1.000]
9	100.0%	1.000	[0.139, 1.000]
10	100.0%	1.000	[0.468, 1.000]
11	100.0%	1.000	[0.665, 1.000]
12	100.0%	1.000	[0.486, 1.000]

**Billingsley, Scheuermann  
& Webber, 2009**

	<b>98.0%</b>	<b>0.959</b>	<b>[0.864, 0.993]</b>
1	100.0%	1.000	[-0.238, 1.000]
2	100.0%	1.000	[-0.127, 1.000]
3	100.0%	1.000	[-0.255, 1.000]
4	80.0%	0.583	[-0.479, 1.000]

5	100.0%	1.000	[-0.238, 1.000]
6	100.0%	1.000	[-0.238, 1.000]
7	100.0%	1.000	[-0.127, 1.000]
8	100.0%	1.000	[-0.127, 1.000]
9	80.0%	0.583	[-0.479, 1.000]
10	100.0%	1.000	[-0.127, 1.000]
11	100.0%	1.000	[-0.127, 1.000]
12	100.0%	1.000	[-0.238, 1.000]
13	80.0%	0.583	[-0.479, 1.000]
14	100.0%	1.000	[-0.238, 1.000]
15	100.0%	1.000	[-0.238, 1.000]
16	100.0%	1.000	[-0.127, 1.000]
17	100.0%	1.000	[-0.238, 1.000]
18	100.0%	1.000	[-0.127, 1.000]
19	100.0%	1.000	[-0.127, 1.000]
20	100.0%	1.000	[-0.127, 1.000]
21	100.0%	1.000	[-0.127, 1.000]
22	100.0%	1.000	[-0.127, 1.000]
23	100.0%	1.000	[-0.399, 1.000]
24	100.0%	1.000	[-0.238, 1.000]
25	100.0%	1.000	[-0.127, 1.000]
26	100.0%	1.000	[-0.127, 1.000]
27	100.0%	1.000	[-0.127, 1.000]

**\*Bell, Young, Blair, &  
Nelson, 1990**

	<b>85.5%</b>	<b>0.709</b>	<b>[0.435, 0.874]</b>
1	100.0%	1.000	[0.242, 1.000]
2	100.0%	1.000	[0.242, 1.000]
3	63.6%	0.267	[-0.436, 0.803]
4	90.9%	0.817	[0.043, 1.000]
5	72.7%	0.411	[-0.294, 0.889]

**Blankenship, Ayres,  
Langone, 2005**

	<b>98.8%</b>	<b>0.973</b>	<b>[0.884, 0.998]</b>
1	100.0%	1.000	[0.153, 1.000]
2	100.0%	1.000	[0.124, 1.000]
3	100.0%	1.000	[0.175, 1.000]
4	100.0%	1.000	[0.153, 1.000]
5	100.0%	1.000	[0.038, 1.000]
6	100.0%	1.000	[0.038, 1.000]

	7	100.0%	1.000	[0.153, 1.000]
	8	100.0%	1.000	[0.153, 1.000]
	9	83.3%	0.556	[-0.146, 0.960]
	10	100.0%	1.000	[0.175, 1.000]
	11	100.0%	1.000	[-0.031, 1.000]
	12	100.0%	1.000	[-0.031, 1.000]
	13	100.0%	1.000	[0.087, 1.000]
	14	100.0%	1.000	[0.087, 1.000]
	15	100.0%	1.000	[0.124, 1.000]
	16	100.0%	1.000	[0.153, 1.000]
	17	100.0%	1.000	[-0.031, 1.000]
	18	100.0%	1.000	[-0.031, 1.000]
<b>Bowman-Perrott, Greenwood, &amp; Tapia, 2007 (study 2 only),</b>				
		<b>80.0%</b>	<b>0.550</b>	<b>[-0.090, 0.916]</b>
<b>Brasch, Williams, &amp; McLaughlin, 2008</b>				
		<b>100.0%</b>	<b>1.000</b>	<b>[0.846, 1.000]</b>
	1	100.0%	1.000	[0.209, 1.000]
	2	100.0%	1.000	[0.335, 1.000]
	3	100.0%	1.000	[0.362, 1.000]
	4	100.0%	1.000	[0.209, 1.000]
	5	100.0%	1.000	[0.371, 1.000]
	6	100.0%	1.000	[0.285, 1.000]
<b>Callicot &amp; Park, 2003</b>				
		<b>72.0%</b>	<b>0.363</b>	<b>[0.262, 0.457]</b>
	1	73.7%	0.321	[-0.181, 0.778]
	2	78.9%	0.367	[-0.141, 0.831]
	3	76.9%	0.350	[-0.215, 0.877]
	4	78.6%	0.533	[-0.126, 0.912]
	5	71.4%	0.500	[-0.231, 0.686]
	6	62.5%	0.200	[-0.252, 0.614]
	7	85.0%	0.700	[0.168, 0.944]
	8	62.5%	0.238	[-0.338, 0.712]
	9	70.0%	0.286	[-0.205, 0.731]
	10	80.0%	0.467	[-0.077, 0.859]
	11	86.7%	0.700	[0.040, 0.973]
	12	84.6%	0.639	[-0.069, 0.968]
	13	73.7%	0.321	[-0.181, 0.778]

14	78.9%	0.367	[-0.141, 0.831]
15	76.9%	0.350	[-0.215, 0.877]
16	78.6%	0.533	[-0.126, 0.912]
17	66.7%	0.500	[-0.231, 0.686]
18	62.5%	0.200	[-0.252, 0.614]
19	50.0%	0.000	[-0.484, 0.484]
20	62.5%	0.238	[-0.338, 0.712]
21	70.0%	0.286	[-0.205, 0.731]
22	75.0%	0.333	[-0.161, 0.782]
23	66.7%	0.250	[-0.301, 0.753]
24	69.2%	0.278	[-0.298, 0.806]

**Cancio, West, & Young,  
2004**

	<b>94.7%</b>	<b>0.892</b>	<b>[0.842, 0.927]</b>
1	86.0%	0.563	[0.190, 0.822]
2	98.1%	0.946	[0.649, 1.000]
3	98.1%	0.921	[0.680, 0.993]
4	100.0%	1.000	[0.719, 1.000]
5	94.3%	0.883	[0.544, 0.990]
6	97.8%	0.949	[0.663, 1.000]
7	84.3%	0.471	[0.085, 0.776]
8	98.1%	0.946	[0.647, 1.000]
9	90.4%	0.803	[0.531, 0.938]
10	100.0%	1.000	[0.719, 1.000]
11	94.3%	0.883	[0.544, 0.990]
12	97.8%	0.949	[0.663, 1.000]

**Carr & Punzo, 1993**

	<b>82.6%</b>	<b>0.636</b>	<b>[0.552, 0.709]</b>
1	100.0%	1.000	[0.596, 1.000]
2	100.0%	1.000	[0.582, 1.000]
3	80.0%	0.375	[-0.090, 0.796]
4	100.0%	1.000	[0.573, 1.000]
5	70.8%	0.378	[-0.102, 0.743]
6	79.2%	0.368	[-0.101, 0.794]
7	100.0%	1.000	[0.586, 1.000]
8	87.5%	0.733	[0.249, 0.951]
9	79.2%	0.368	[-0.101, 0.794]
10	100.0%	1.000	[0.596, 1.000]
11	64.0%	0.219	[-0.225, 0.624]

	12	80.0%	0.375	[-0.090, 0.796]
	13	100.0%	1.000	[0.573, 1.000]
	14	62.5%	1.000	[0.573, 1.000]
	15	83.3%	0.495	[-0.022, 0.867]
	16	58.3%	0.143	[-0.309, 0.563]
	17	62.5%	1.000	[0.573, 1.000]
	18	79.2%	0.368	[-0.101, 0.794]
<b>Cieslar, McLaughlin, &amp; Derby, 2008</b>		<b>100.0%</b>	<b>1.000</b>	<b>[0.841, 1.000]</b>
	1	100.0%	1.000	[0.174, 1.000]
	2	100.0%	1.000	[0.242, 1.000]
	3	100.0%	1.000	[0.153, 1.000]
	4	100.0%	1.000	[0.175, 1.000]
	5	100.0%	1.000	[0.280, 1.000]
	6	100.0%	1.000	[0.292, 1.000]
<b>Daly, Garbacz, Olson, Persampieri &amp; Ni, 2006</b>		<b>94.4%</b>	<b>0.874</b>	<b>[0.742, 0.946]</b>
	1	50.0%	0.000	[-0.806, 0.806]
	2	100.0%	1.000	[-0.031, 1.000]
	3	100.0%	1.000	[0.087, 1.000]
	4	91.7%	0.778	[-0.007, 1.000]
	5	100.0%	1.000	[-0.127, 1.000]
	6	87.5%	0.733	[-0.188, 1.000]
	7	90.0%	0.762	[-0.074, 1.000]
	8	100.0%	1.000	[0.194, 1.000]
	9	66.7%	0.333	[-0.625, 0.940]
	10	100.0%	1.000	[-0.031, 1.000]
	11	100.0%	1.000	[0.087, 1.000]
	12	100.0%	1.000	[0.175, 1.000]
	13	100.0%	1.000	[-0.127, 1.000]
	14	100.0%	1.000	[0.038, 1.000]
	15	100.0%	1.000	[0.124, 1.000]
	16	100.0%	1.000	[0.194, 1.000]
<b>Ducharme &amp; Harris, 2005</b>		<b>92.9%</b>	<b>0.819</b>	<b>[0.729, 0.883]</b>
	1	63.6%	0.273	[-0.225, 0.673]
	2	89.3%	0.781	[0.360, 0.960]
	3	91.1%	0.592	[0.178, 0.869]

	4	87.5%	0.609	[0.282, 0.829]
	5	100.0%	1.000	[0.564, 1.000]
	6	100.0%	1.000	[0.640, 1.000]
	7	100.0%	1.000	[0.610, 1.000]
	8	100.0%	1.000	[0.774, 1.000]
<b>Dunlap, White, Vera, Wilson, &amp; Panace, 1996</b>		<b>100.0%</b>	<b>1.000</b>	<b>[0.884, 1.000]</b>
	1	100.0%	1.000	[0.427, 1.000]
	2	100.0%	1.000	[0.485, 1.000]
	3	100.0%	1.000	[0.331, 1.000]
	4	100.0%	1.000	[0.427, 1.000]
	5	100.0%	1.000	[0.485, 1.000]
	6	100.0%	1.000	[0.331, 1.000]
<b>DuPaul, McGoey, &amp; Yugar, 1997</b>		<b>82.1%</b>	<b>0.629</b>	<b>[0.504, 0.729]</b>
	1	92.9%	0.854	[0.191, 1.000]
	2	73.7%	0.321	[-0.181, 0.778]
	3	90.9%	0.771	[-0.037, 1.000]
	4	60.9%	0.179	[-0.283, 0.603]
	5	88.5%	0.729	[0.246, 0.950]
	6	77.8%	0.357	[-0.158, 0.829]
	7	78.6%	0.563	[-0.102, 0.918]
	8	80.0%	0.524	[-0.028, 0.875]
	9	81.8%	0.542	[-0.181, 0.959]
	10	73.9%	0.452	[-0.045, 0.797]
	11	96.2%	0.910	[0.452, 1.000]
	12	94.4%	0.839	[0.175, 1.000]
<b>Falk &amp; Wehby, 2001</b>		<b>86.4%</b>	<b>0.656</b>	<b>[0.292, 0.874]</b>
	1	83.3%	0.625	[-0.339, 1.000]
	2	100.0%	1.000	[-0.018, 1.000]
	3	71.4%	0.300	[-0.361, 0.937]
	4	87.5%	0.667	[-0.194, 1.000]
	5	85.7%	0.650	[-0.252, 1.000]
	6	87.5%	0.667	[-0.194, 1.000]
<b>Ferguson, Ashbaugh, O'Reilly, &amp; McLaughlin,</b>		<b>83.8%</b>	<b>0.655</b>	<b>[0.414, 0.819]</b>



<b>2004</b>				
	1	66.7%	0.148	[-0.298, 0.716]
	2	76.2%	0.495	[-0.032, 0.837]
	3	87.5%	0.667	[0.008, 0.970]
	4	100.0%	1.000	[0.432, 1.000]
<b>Foxx, Kyle, Faw, &amp; Bittle, 1989</b>		<b>100.0%</b>	<b>1.000</b>	<b>[0.655, 1.000]</b>
	1	100.0%	1.000	[-0.127, 1.000]
	2	100.0%	1.000	[0.087, 1.000]
	3	100.0%	1.000	[-0.127, 1.000]
	4	100.0%	1.000	[-0.087, 1.000]
<b>Freeman &amp; Dexter-Mazza, 2004</b>		<b>63.3%</b>	<b>0.236</b>	<b>[-0.178, 0.598]</b>
	1	68.4%	0.269	[-0.232, 0.724]
	2	54.5%	0.083	[-0.564, 0.694]
<b>Gulchak, 2008</b>		<b>52.6%</b>	<b>0.050</b>	<b>[-0.453, 0.536]</b>
<b>Gunter, Miller, &amp; Venn, 2003</b>		<b>75.6%</b>	<b>0.379</b>	<b>[0.009, 0.690]</b>
<b>Hale, Skinner, Winn, Oliver, Allin, &amp; Molloy, 2005</b>		<b>78.9%</b>	<b>0.536</b>	<b>[0.377, 0.668]</b>
	1	75.0%	0.467	[-0.363, 0.952]
	2	64.3%	0.458	[0.042, 0.558]
	3	66.7%	0.250	[-0.347, 0.799]
	4	77.8%	0.500	[-0.283, 0.955]
	5	78.6%	0.533	[-0.126, 0.912]
	6	66.7%	0.250	[-0.347, 0.799]
	7	100.0%	1.000	[0.038, 1.000]
	8	78.6%	0.533	[-0.126, 0.912]
	9	71.4%	0.378	[-0.245, 0.834]
	10	75.0%	0.438	[-0.239, 0.894]
	11	100.0%	1.000	[0.087, 1.000]
	12	100.0%	1.000	[0.335, 1.000]
	13	78.6%	0.533	[-0.126, 0.912]
	14	83.3%	0.625	[-0.106, 0.967]
		<b>70.3%</b>	<b>0.405</b>	<b>[0.228, 0.558]</b>

**Hansen & Lignagarist-Kraft, 2005**

1	57.1%	0.143	[-0.464, 0.673]
2	50.0%	0.000	[-0.572, 0.572]
3	100.0%	1.000	[0.335, 1.000]
4	100.0%	1.000	[0.242, 1.000]
5	94.4%	0.889	[0.344, 1.000]
6	57.1%	0.143	[-0.464, 0.673]
7	100.0%	1.000	[0.371, 1.000]
8	35.7%	-0.400	[-0.556, 0.274]
9	54.5%	0.083	[-0.564, 0.694]
10	50.0%	0.000	[-0.572, 0.572]

**Hawkins & Axelrod, 2008**

	<b>77.8%</b>	<b>0.556</b>	<b>[0.308, 0.739]</b>
1	75.0%	0.500	[-0.389, 0.955]
2	100.0%	1.000	[0.058, 1.000]
3	100.0%	1.000	[0.058, 1.000]
4	50.0%	0.000	[-0.726, 0.726]
5	75.0%	0.500	[-0.389, 0.955]
6	87.5%	0.750	[-0.181, 1.000]
7	50.0%	0.000	[-0.726, 0.726]
8	62.5%	0.500	[-0.389, 0.955]
9	100.0%	1.000	[0.058, 1.000]

**Hughes, Deshler, Ruhl, & Schumaker, 1993**

	<b>100.0%</b>	<b>1.000</b>	<b>[0.669, 1.000]</b>
1	100.0%	1.000	[-0.018, 1.000]
2	100.0%	1.000	[-0.018, 1.000]
3	100.0%	1.000	[0.087, 1.000]
4	100.0%	1.000	[0.175, 1.000]

**Jolivette, Wehby, Canale, & Massey, 2001**

	<b>84.4%</b>	<b>0.477</b>	<b>[0.346, 0.594]</b>
1	71.4%	0.300	[-0.182, 0.736]
2	100.0%	1.000	[0.279, 1.000]
3	87.5%	0.550	[0.000, 0.916]
4	81.0%	0.533	[-0.009, 0.877]
5	100.0%	1.000	[0.279, 1.000]
6	87.5%	0.550	[0.000, 0.916]
7	90.5%	0.767	[0.206, 0.979]
8	78.3%	0.042	[-0.150, 0.676]

	9	95.8%	0.850	[0.226, 1.000]
	10	61.9%	0.067	[-0.311, 0.570]
	11	78.3%	0.042	[-0.150, 0.676]
	12	83.3%	0.400	[-0.081, 0.840]
	13	71.4%	0.300	[-0.182, 0.736]
	14	87.0%	0.425	[-0.076, 0.891]
	15	87.5%	0.550	[0.000, 0.916]
<b>Kern, Childs, Dunlap, Clarke, &amp; Falk, 1994</b>		<b>95.5%</b>	<b>0.907</b>	<b>[0.743, 0.976]</b>
	1	100.0%	1.000	[0.465, 1.000]
	2	100.0%	1.000	[0.586, 1.000]
	3	88.9%	0.766	[0.402, 0.939]
<b>Kern, Delaney, Clarke, Dunlap, &amp; Childs, 2001</b>		<b>69.6%</b>	<b>0.379</b>	<b>[0.246, 0.499]</b>
	1	100.0%	1.000	[0.582, 1.000]
	2	64.0%	0.219	[-0.225, 0.624]
	3	68.0%	0.306	[-0.154, 0.688]
	4	70.4%	0.407	[-0.044, 0.736]
	5	70.4%	0.407	[-0.044, 0.736]
	6	55.6%	0.110	[-0.321, 0.509]
	7	63.0%	0.258	[-0.186, 0.627]
	8	59.3%	0.156	[-0.270, 0.552]
	9	77.8%	0.540	[0.085, 0.830]
<b>Kern, Dunlap, Childs &amp; Clarke, 1994</b>		<b>70.1%</b>	<b>0.383</b>	<b>[0.295, 0.465]</b>
	1	67.6%	0.546	[0.237, 0.767]
	2	88.1%	0.700	[0.399, 0.879]
	3	80.0%	0.546	[0.237, 0.767]
	4	68.6%	0.403	[0.106, 0.645]
	5	63.6%	0.100	[-0.227, 0.413]
	6	55.9%	0.105	[-0.273, 0.464]
	7	72.9%	0.314	[0.014, 0.587]
	8	72.7%	0.403	[0.097, 0.652]
	9	58.8%	-0.214	[-0.470, 0.101]
	10	63.6%	0.100	[-0.227, 0.413]
<b>Kern, Wacker, Mace, Falk,</b>		<b>71.7%</b>	<b>0.433</b>	<b>[0.326, 0.531]</b>

**Dunlap, & Kromery, 1995  
(experiment 1)**

	54.5%	0.083	[-0.564, 0.694]
2	71.4%	0.300	[0.259, 0.813]
3	57.9%	0.136	[0.369, 0.606]
4	66.7%	0.000	[-0.613, 0.613]
5	55.6%	0.100	[-0.419, 0.588]
6	64.7%	0.220	[-0.299, 0.707]
v7	76.9%	0.350	[-0.100, 0.756]
8	74.2%	0.410	[-0.015, 0.737]
9	77.8%	0.532	[-0.044, 0.877]
10	54.5%	0.083	[-0.564, 0.694]
11	71.4%	0.300	[0.259, 0.813]
12	63.2%	0.244	[-0.283, 0.684]
13	50.0%	0.000	[-0.613, 0.613]
14	55.6%	0.100	[-0.419, 0.588]
15	70.6%	0.356	[-0.208, 0.790]
16	80.8%	0.458	[-0.026, 0.824]
17	96.8%	0.926	[0.536, 1.000]
18	100.0%	1.000	[0.465, 1.000]

**Kern, Wacker, Mace, Falk,  
Dunlap, & Kromery, 1994  
(experiment 2)**

	<b>80.2%</b>	<b>0.580</b>	<b>[0.490, 0.659]</b>
1	73.9%	0.425	[-0.072, 0.786]
2	57.1%	0.091	[-0.363, 0.557]
3	90.9%	0.771	[-0.037, 1.000]
4	83.3%	0.625	[-0.106, 0.967]
5	91.3%	0.808	[0.310, 0.983]
6	76.2%	0.495	[-0.032, 0.837]
7	100.0%	1.000	[0.153, 1.000]
8	100.0%	1.000	[0.242, 1.000]
9	52.6%	0.050	[-0.453, 0.536]
10	85.7%	0.714	[0.048, 0.975]
11	81.8%	0.657	[-0.076, 0.969]
12	66.7%	0.250	[-0.263, 0.716]
13	73.3%	0.300	[-0.259, 0.813]
14	90.9%	0.771	[-0.037, 1.000]
15	69.2%	0.278	[-0.298, 0.806]
16	88.9%	0.679	[-0.152, 1.000]

17	85.7%	0.417	[-0.135, 0.948]
18	91.7%	0.700	[-0.076, 1.000]
19	88.9%	0.438	[-0.125, 1.000]
20	52.6%	0.050	[-0.453, 0.536]
21	92.9%	0.857	[0.201, 1.000]
22	90.9%	0.804	[0.015, 1.000]
23	72.2%	0.375	[-0.175, 0.796]
24	86.7%	0.700	[0.040, 0.973]
25	81.8%	0.542	[-0.181, 0.959]
26	69.2%	0.278	[0.298, 0.806]
27	100.0%	1.000	[0.016, 1.000]
28	100.0%	1.000	[0.103, 1.000]
29	91.7%	0.700	[-0.076, 1.000]
30	88.9%	0.438	[-0.125, 1.000]

**Kern-Dunlap, Dunlap,  
Clarke, & Childs, 1992**

	<b>78.0%</b>	<b>0.554</b>	<b>[0.417, 0.668]</b>
1	82.9%	0.606	[0.238, 0.840]
2	71.4%	0.417	[-0.104, 0.783]
3	56.3%	0.125	[-0.440, 0.630]
4	72.7%	0.375	[-0.154, 0.766]
5	85.4%	0.606	[0.238, 0.840]
6	90.5%	0.806	[0.297, 0.983]
7	87.5%	0.750	[0.140, 0.978]
8	63.6%	0.214	[-0.256, 0.644]

**Lane, Eisner, Kretzer,  
Bruhn, Crnabori, Funke, et  
al., 2009**

**88.9%**      **0.778**      **[0.216, 0.980]**

**Lawson & Greer  
(experiment 1)**

	<b>93.1%</b>	<b>0.862</b>	<b>[0.718, 0.941]</b>
1	71.4%	0.417	[-0.477, 0.948]
2	88.9%	0.775	[-0.098, 1.000]
3	91.7%	0.833	[0.104, 1.000]
4	100.0%	1.000	[-0.031, 1.000]
5	100.0%	1.000	[0.125, 1.000]
6	66.7%	0.333	[-0.361, 0.822]
7	100.0%	1.000	[0.087, 1.000]
8	100.0%	1.000	[0.192, 1.000]

	9	100.0%	1.000	[0.292, 1.000]
	10	100.0%	1.000	[0.087, 1.000]
	11	100.0%	1.000	[0.192, 1.000]
	12	100.0%	1.000	[0.175, 1.000]
<b>Lawson &amp; Greer, 2006</b>				
<b>(experiment 2)</b>				
		<b>89.1%</b>	<b>0.778</b>	<b>[0.538, 0.996]</b>
	1	85.7%	0.708	[-0.279, 1.000]
	2	87.5%	0.750	[-0.181, 1.000]
	3	87.5%	0.733	[-0.188, 1.000]
	4	88.9%	0.750	[-0.122, 1.000]
	5	85.7%	0.708	[-0.279, 1.000]
	6	87.5%	0.750	[-0.181, 1.000]
	7	87.5%	0.733	[-0.188, 1.000]
	8	100.0%	1.000	[0.087, 1.000]
<b>Lloyd &amp; Hilliard, 1989</b>				
		<b>80.6%</b>	<b>0.611</b>	<b>[0.421, 0.753]</b>
	1	52.0%	0.840	[0.400, 0.986]
	2	94.3%	0.266	[-0.117, 0.600]
	3	78.6%	0.856	[0.454, 0.987]
	4	95.0%	0.890	[0.355, 1.000]
<b>Locke &amp; Fuchs, 1995</b>				
		<b>87.2%</b>	<b>0.739</b>	<b>[0.523, 0.874]</b>
	1	90.0%	0.798	[0.274, 0.982]
	2	89.5%	0.784	[0.236, 0.981]
	3	90.0%	0.798	[0.274, 0.982]
	4	78.9%	0.568	[0.010, 0.887]
<b>Maher, 1986</b>				
		<b>100.0%</b>	<b>1.000</b>	<b>[0.851, 1.000]</b>
	1	100.0%	1.000	[0.192, 1.000]
	2	100.0%	1.000	[0.354, 1.000]
	3	100.0%	1.000	[0.192, 1.000]
	4	100.0%	1.000	[0.354, 1.000]
	5	100.0%	1.000	[0.192, 1.000]
	6	100.0%	1.000	[0.354, 1.000]
<b>McCurdy, Cundari, &amp; Lenz, 1990</b>				
		<b>97.6%</b>	<b>0.951</b>	<b>[0.796, 0.996]</b>
	1	100.0%	1.000	[-0.132, 1.000]
	2	100.0%	1.000	[-0.132, 1.000]

	3	93.3%	0.850	[0.187, 1.000]
	4	93.3%	0.850	[0.187, 1.000]
	5	100.0%	1.000	[-0.132, 1.000]
	6	100.0%	1.000	[-0.132, 1.000]
	7	100.0%	1.000	[0.354, 1.000]
	8	100.0%	1.000	[0.354, 1.000]
<b>McCurdy &amp; Shapiro, 1988</b>		<b>78.1%</b>	<b>0.551</b>	<b>[0.408, 0.669]</b>
	1	66.7%	0.306	[-0.293, 0.772]
	2	72.7%	0.436	[-0.072, 0.790]
	3	72.4%	0.492	[-0.119, 0.866]
	4	75.0%	0.492	[-0.119, 0.866]
	5	90.9%	0.812	[0.317, 0.983]
	6	89.3%	0.754	[0.301, 0.955]
	7	76.7%	0.514	[0.086, 0.802]
	8	76.0%	0.513	[0.040, 0.820]
<b>McKenney &amp; Dattilo, 2001</b>		<b>80.4%</b>	<b>0.586</b>	<b>[0.426, 0.714]</b>
	1	95.2%	0.869	[0.277, 1.000]
	2	85.7%	0.708	[0.189, 0.946]
	3	71.4%	0.300	[-0.182, 0.736]
	4	66.7%	0.325	[-0.458, 0.872]
	5	66.7%	0.500	[-0.301, 0.753]
	6	95.2%	0.868	[0.272, 1.000]
	7	55.6%	0.100	[-0.598, 0.756]
	8	80.0%	0.550	[-0.090, 0.916]
	9	81.0%	0.475	[-0.060, 0.861]
<b>McLaughlin, 1992</b>		<b>98.5%</b>	<b>0.971</b>	<b>[0.916, 0.993]</b>
	1	98.2%	0.956	[0.706, 1.000]
	2	100.0%	1.000	[0.797, 1.000]
	3	100.0%	1.000	[0.807, 1.000]
	4	94.5%	0.889	[0.650, 0.980]
	5	100.0%	1.000	[0.789, 1.000]
<b>Miller &amp; Cole, 1998</b>		<b>90.9%</b>	<b>0.818</b>	<b>[0.594, 0.934]</b>
	1	100.0%	1.000	[0.479, 1.000]
	2	95.5%	0.908	[0.440, 1.000]
	3	77.3%	0.353	[-0.127, 0.788]
		<b>93.1%</b>	<b>0.856</b>	<b>[0.816, 0.888]</b>

**Moore, Cartledge, &  
Heckaman, 1995**

1	100.0%	1.000	[0.448, 1.000]
2	76.0%	0.449	[-0.031, 0.795]
3	90.9%	0.752	[0.305, 0.954]
4	82.4%	0.536	[-0.155, 0.913]
5	88.0%	0.724	[0.234, 0.949]
6	87.9%	0.670	[0.221, 0.914]
7	68.8%	0.365	[-0.232, 0.793]
8	100.0%	1.000	[0.521, 1.000]
9	80.0%	0.441	[-0.008, 0.792]
10	100.0%	1.000	[0.427, 1.000]
11	91.3%	0.795	[0.276, 0.982]
12	86.7%	0.620	[0.150, 0.903]
13	93.3%	0.866	[0.240, 1.000]
14	95.2%	0.883	[0.330, 1.000]
15	92.6%	0.786	[0.265, 0.981]
16	86.7%	0.732	[0.094, 0.976]
17	85.7%	0.650	[0.093, 0.935]
18	81.5%	0.458	[-0.026, 0.824]
19	100.0%	1.000	[0.646, 1.000]
20	87.5%	0.749	[0.363, 0.935]
21	100.0%	1.000	[0.568, 1.000]
22	81.8%	0.570	[0.147, 0.842]
23	93.8%	0.875	[0.514, 0.989]
24	100.0%	1.000	[0.568, 1.000]
25	96.7%	0.921	[0.508, 1.000]
26	100.0%	1.000	[0.665, 1.000]
27	93.3%	0.814	[0.336, 0.983]
28	100.0%	1.000	[0.614, 1.000]
29	100.0%	1.000	[0.665, 1.000]
30	100.0%	1.000	[0.560, 1.000]
31	100.0%	1.000	[0.569, 1.000]
32	100.0%	1.000	[0.630, 1.000]
33	100.0%	1.000	[0.509, 1.000]
34	100.0%	1.000	[0.569, 1.000]
35	100.0%	1.000	[0.630, 1.000]
36	100.0%	1.000	[0.509, 1.000]
	<b>100.0%</b>	<b>1.000</b>	<b>[0.693, 1.000]</b>



**Mulcahy, Krezmien,  
Leone, Houchins, &  
Baltodano, 2008**

1	100.0%	1.000	[-0.238, 1.000]
2	100.0%	1.000	[-0.177, 1.000]
3	100.0%	1.000	[-0.238, 1.000]
4	100.0%	1.000	[-0.031, 1.000]
5	100.0%	1.000	[-0.238, 1.000]
6	100.0%	1.000	[-0.031, 1.000]

**Musser, Bray, Kehle, &  
Jenson, 2001**

	<b>98.1%</b>	<b>0.958</b>	<b>[0.721, 1.000]</b>
1	100.0%	1.000	[0.292, 1.000]
2	100.0%	1.000	[0.438, 1.000]
3	95.8%	0.889	[0.359, 1.000]

**Nelson & Behler, 1989**

	<b>77.3%</b>	<b>0.509</b>	<b>[0.163, 0.758]</b>
1	75.0%	0.598	[-0.157, 0.810]
2	75.0%	0.438	[-0.239, 0.894]
3	90.0%	0.792	[-0.034, 1.000]
4	70.0%	0.375	[-0.365, 0.882]

**Ninness, Ellis, Miller, &  
Baker, 1995**

	<b>100.0%</b>	<b>1.000</b>	<b>[0.854, 1.000]</b>
1	100.0%	1.000	[0.552, 1.000]
2	100.0%	1.000	[0.448, 1.000]
3	100.0%	1.000	[0.552, 1.000]
4	100.0%	1.000	[0.682, 1.000]

**Ninness, Fuerst, &  
Rutherford, 1995**

	<b>98.5%</b>	<b>0.970</b>	<b>[0.793, 1.000]</b>
1	100.0%	1.000	[-0.031, 1.000]
2	85.7%	0.708	[-0.279, 1.000]
3	100.0%	1.000	[0.242, 1.000]
4	100.0%	1.000	[0.242, 1.000]
5	100.0%	1.000	[0.175, 1.000]
6	100.0%	1.000	[0.326, 1.000]

**Ogier & Hornby, 1996**

	<b>90.0%</b>	<b>0.800</b>	<b>[0.647, 0.896]</b>
1	100.0%	1.000	[0.669, 1.000]

	2	97.5%	0.950	[0.670, 1.000]
	3	72.5%	0.273	[-0.078, 0.617]
<b>Patterson, 2005 (study 2 only)</b>		<b>100.0%</b>	<b>1.000</b>	<b>[0.787, 1.000]</b>
	1	100.0%	1.000	[0.639, 1.000]
	2	100.0%	1.000	[0.546, 1.000]
<b>Popkin &amp; Skinner, 2003</b>		<b>80.0%</b>	<b>0.593</b>	<b>[0.377, 0.753]</b>
	1	93.3%	0.841	[0.411, 0.986]
	2	66.7%	0.306	[-0.115, 0.652]
	3	80.0%	0.375	[-0.059, 0.766]
<b>Prater, Hogan, &amp; Miller, 1992</b>		<b>100.0%</b>	<b>1.000</b>	<b>[0.933, 1.000]</b>
	1	100.0%	1.000	[0.058, 1.000]
	2	100.0%	1.000	[-0.031, 1.000]
	3	100.0%	1.000	[0.058, 1.000]
	4	100.0%	1.000	[-0.031, 1.000]
	5	100.0%	1.000	[0.315, 1.000]
	6	100.0%	1.000	[0.693, 1.000]
	7	100.0%	1.000	[0.571, 1.000]
	8	100.0%	1.000	[0.310, 1.000]
	9	100.0%	1.000	[0.310, 1.000]
	10	100.0%	1.000	[-0.410, 1.000]
	11	100.0%	1.000	[-0.410, 1.000]
	12	100.0%	1.000	[-0.410, 1.000]
	13	100.0%	1.000	[-0.238, 1.000]
	14	100.0%	1.000	[-0.238, 1.000]
	15	100.0%	1.000	[-0.064, 1.000]
	16	100.0%	1.000	[-0.064, 1.000]
<b>Rafferty &amp; Raimondi, 2009</b>		<b>92.5%</b>	<b>0.831</b>	<b>[0.717, 0.905]</b>
	1	70.0%	0.375	[-0.365, 0.882]
	2	100.0%	1.000	[0.242, 1.000]
	3	100.0%	1.000	[0.087, 1.000]
	4	100.0%	1.000	[0.266, 1.000]
	5	92.3%	0.819	[0.084, 1.000]
	6	100.0%	1.000	[0.062, 1.000]
	7	70.0%	0.375	[-0.365, 1.000]

8	100.0%	1.000	[0.242, 1.000]
9	100.0%	1.000	[0.087, 1.000]
10	84.6%	0.639	[-0.069, 0.968]
11	76.9%	0.458	[-0.196, 0.898]
12	100.0%	1.000	[0.062, 1.000]
13	100.0%	1.000	[0.175, 1.000]
14	100.0%	1.000	[0.242, 1.000]
15	100.0%	1.000	[0.087, 1.000]
16	92.3%	0.819	[0.084, 1.000]
17	84.6%	0.639	[-0.069, 0.968]
18	100.0%	1.000	[0.062, 1.000]
<b>Rasmussen &amp; O'Neill, 2006</b>	<b>91.9%</b>	<b>0.832</b>	<b>[0.685, 0.920]</b>
1	88.9%	0.770	[0.333, 0.957]
2	76.0%	0.513	[0.040, 0.820]
3	95.7%	0.912	[0.457, 1.000]
4	100.0%	1.000	[0.112, 1.000]
5	100.0%	1.000	[0.397, 1.000]
6	100.0%	1.000	[0.233, 1.000]
<b>Salend, Reynolds, &amp; Coyle, 1989</b>	<b>99.0%</b>	<b>0.979</b>	<b>[0.911, 0.998]</b>
1	100.0%	1.000	[0.631, 1.000]
2	100.0%	1.000	[0.631, 1.000]
3	96.4%	0.927	[0.539, 1.000]
4	100.0%	1.000	[0.640, 1.000]
5	96.6%	0.929	[0.550, 1.000]
6	100.0%	1.000	[0.648, 1.000]
7	100.0%	1.000	[0.648, 1.000]
<b>Salend &amp; Sonnenschein, 1989</b>	<b>99.5%</b>	<b>0.989</b>	<b>[0.920, 1.000]</b>
1	100.0%	1.000	[0.331, 1.000]
2	100.0%	1.000	[0.448, 1.000]
3	100.0%	1.000	[0.362, 1.000]
4	94.1%	0.879	[0.298, 1.000]
5	100.0%	1.000	[0.331, 1.000]
6	100.0%	1.000	[0.448, 1.000]
7	100.0%	1.000	[0.331, 1.000]
8	100.0%	1.000	[0.448, 1.000]

	9	100.0%	1.000	[0.362, 1.000]
	10	100.0%	1.000	[0.448, 1.000]
	11	100.0%	1.000	[0.331, 1.000]
	12	100.0%	1.000	[0.448, 1.000]
<b>Schoenfeld &amp; Mathur, 2009</b>		<b>88.0%</b>	<b>0.760</b>	<b>[0.543, 0.890]</b>
	1	93.8%	0.855	[0.208, 1.000]
	2	76.5%	0.528	[-0.064, 0.876]
	3	81.0%	0.596	[0.065, 0.894]
	4	100.0%	1.000	[-0.132, 1.000]
	5	100.0%	1.000	[-0.031, 1.000]
	6	100.0%	1.000	[0.058, 1.000]
<b>Singh, Lancioni, Singh Joy, Winton, Sabaawi, Wahler, et al., 2007</b>		<b>76.4%</b>	<b>0.527</b>	<b>[0.358, 0.665]</b>
	1	50.0%	0.000	[-0.726, 0.726]
	2	72.7%	0.411	[-0.294, 0.889]
	3	73.3%	0.318	[-0.227, 0.818]
	4	50.0%	0.000	[-0.726, 0.726]
	5	63.6%	0.214	[-0.408, 0.788]
	6	73.3%	0.318	[-0.227, 0.818]
	7	94.4%	0.869	[0.499, 0.988]
	8	77.8%	0.476	[0.078, 0.768]
<b>Skinner, Belfiore, &amp; Pierce, 1992</b>		<b>100.0%</b>	<b>1.000</b>	<b>[0.876, 1.000]</b>
	1	100.0%	1.000	[-0.031, 1.000]
	2	100.0%	1.000	[0.310, 1.000]
	3	100.0%	1.000	[0.539, 1.000]
	4	100.0%	1.000	[0.586, 1.000]
	5	100.0%	1.000	[0.479, 1.000]
<b>Skinner, Hurst, Teeple, &amp; Meadows, 2002</b>		<b>74.5%</b>	<b>0.489</b>	<b>[0.165, 0.728]</b>
	1	63.6%	0.260	[-0.436, 0.803]
	2	85.7%	0.714	[0.040, 0.975]
	3	54.5%	0.083	[-0.564, 0.694]
	4	90.9%	0.817	[0.043, 1.000]
		<b>94.0%</b>	<b>0.559</b>	<b>[0.407, 0.687]</b>

**Skinner & Johnson, 1995**

1	98.4%	0.920	[0.514, 1.000]
2	98.4%	0.920	[0.514, 1.000]
3	89.1%	0.439	[0.071, 0.769]
4	95.3%	0.759	[0.339, 0.955]
5	93.8%	0.679	[0.262, 0.916]
6	89.1%	0.439	[0.071, 0.769]

**Smith (B), & Sugai, & Brown, 2000**

	<b>66.7%</b>	<b>0.333</b>	<b>[0.009, 0.601]</b>
1	56.3%	0.125	[-0.440, 0.630]
2	50.0%	0.000	[-0.538, 0.538]
3	93.8%	0.875	[0.280, 1.000]

**Smith (D), Nelson, Young, & West, 1992**

	<b>90.4%</b>	<b>0.748</b>	<b>[0.680, 0.805]</b>
1	100.0%	1.000	[0.153, 1.000]
2	100.0%	1.000	[-0.005, 1.000]
3	100.0%	1.000	[0.415, 1.000]
4	88.9%	0.679	[-0.152, 1.000]
5	94.7%	0.892	[0.361, 1.000]
6	100.0%	1.000	[0.250, 1.000]
7	95.5%	0.902	[0.410, 1.000]
8	92.9%	0.857	[0.201, 1.000]
9	79.3%	0.574	[0.140, 0.843]
10	89.7%	0.442	[-0.043, 0.895]
11	100.0%	1.000	[0.293, 1.000]
12	86.4%	0.421	[-0.083, 0.891]
13	92.9%	0.854	[0.191, 1.000]
14	94.4%	0.800	[0.088, 1.000]
15	100.0%	1.000	[0.242, 1.000]
16	80.0%	0.524	[-0.225, 0.957]
17	83.3%	0.625	[-0.106, 0.967]
18	80.0%	0.467	[-0.077, 0.859]
19	94.7%	0.864	[0.255, 1.000]
20	92.9%	0.857	[0.201, 1.000]
21	100.0%	1.000	[0.263, 1.000]
22	89.7%	0.442	[-0.043, 0.895]
23	100.0%	1.000	[0.164, 1.000]
24	86.4%	0.421	[-0.083, 0.891]

	25	93.3%	0.866	[0.240, 1.000]
	26	71.4%	0.300	[-0.182, 0.736]
	27	95.7%	0.909	[0.444, 1.000]
	28	100.0%	1.000	[0.402, 1.000]
	29	88.0%	0.724	[0.234, 0.949]
	30	76.5%	0.346	[-0.055, 0.707]
	31	93.8%	0.833	[0.391, 0.985]
	32	73.1%	0.316	[-0.127, 0.715]
<b>Sprute, Williams, &amp; McLaughlin, 1990</b>				
		<b>96.7%</b>	<b>0.933</b>	<b>[0.726, 0.994]</b>
<b>Strong, Wehby, Falk, &amp; Lane, 2004</b>				
		<b>69.0%</b>	<b>0.313</b>	<b>[0.017, 0.572]</b>
	1	55.0%	-0.200	[-0.333, 0.388]
	2	77.8%	0.550	[-0.291, 0.960]
	3	75.0%	0.333	[-0.161, 0.782]
	4	77.8%	0.550	[-0.291, 0.960]
<b>Sutherland &amp; Snyder, 2007</b>				
		<b>79.0%</b>	<b>0.567</b>	<b>[0.471, 0.650]</b>
	1	97.0%	0.938	[0.60, 1.000]
	2	73.5%	0.463	[0.068, 0.748]
	3	82.5%	0.627	[0.268, 0.849]
	4	93.2%	0.790	[0.395, 0.961]
	5	77.4%	0.544	[0.129, 0.814]
	6	63.6%	0.267	[-0.132, 0.601]
	7	65.0%	0.253	[-0.103, 0.569]
	8	88.6%	0.651	[0.249, 0.889]
	9	40.0%	-0.250	[-0.644, 0.504]
	10	72.7%	0.450	[-0.293, 0.896]
	11	60.0%	0.167	[-0.490, 0.775]
	12	80.0%	0.375	[-0.216, 0.944]
	13	50.0%	-0.042	[-0.584, 0.648]
	14	100.0%	1.000	[0.242, 1.000]
	15	80.0%	0.583	[-0.213, 0.963]
	16	100.0%	1.000	[0.042, 1.000]
<b>Swain &amp; McLaughlin, 1998</b>				
		<b>73.4%</b>	<b>0.420</b>	<b>[0.271, 0.551]</b>
	1	100.0%	1.000	[0.524, 1.000]
	2	78.0%	0.073	[-0.358, 0.156]

	3	77.0%	0.429	[0.128, 0.676]
	4	55.9%	0.118	[-0.167, 0.388]
<b>Swanson, 1985</b>		<b>94.0%</b>	<b>0.826</b>	<b>[0.698, 0.907]</b>
	1	75.0%	0.333	[-0.297, 0.940]
	2	75.0%	-0.333	[-0.297, 0.940]
	3	84.6%	0.409	[-0.149, 0.947]
	4	84.6%	0.409	[-0.149, 0.464]
	5	100.0%	1.000	[-0.018, 1.000]
	6	100.0%	1.000	[-0.018, 1.000]
	7	100.0%	1.000	[0.091, 1.000]
	8	100.0%	1.000	[0.091, 1.000]
	9	75.0%	0.333	[-0.297, 0.940]
	10	100.0%	1.000	[-0.018, 1.000]
	11	100.0%	1.000	[-0.399, 1.000]
	12	75.0%	0.500	[-0.700, 1.000]
	13	100.0%	1.000	[-0.018, 1.000]
	14	87.5%	0.667	[-0.194, 1.000]
	15	100.0%	1.000	[0.091, 1.000]
	16	92.3%	0.705	[-0.060, 1.000]
	17	100.0%	1.000	[-0.018, 1.000]
	18	100.0%	1.000	[-0.018, 1.000]
	19	100.0%	1.000	[0.091, 1.000]
	20	100.0%	1.000	[0.091, 1.000]
	21	100.0%	1.000	[-0.018, 1.000]
	22	100.0%	1.000	[-0.018, 1.000]
	23	100.0%	1.000	[-0.399, 1.000]
	24	100.0%	1.000	[-0.399, 1.000]
<b>Theodore, Bray, &amp; Kehle, 2004</b>		<b>100.0%</b>	<b>1.000</b>	<b>[0.838, 1.000]</b>
	1	100.0%	1.000	[0.661, 1.000]
	2	100.0%	1.000	[0.655, 1.000]
	3	100.0%	1.000	[0.661, 1.000]
	4	100.0%	1.000	[0.586, 1.000]
	5	100.0%	1.000	[0.331, 1.000]
<b>Truchlicka, McLaughlin &amp; Swain, 1998</b>		<b>69.1%</b>	<b>0.380</b>	<b>[0.178, 0.554]</b>
	1	70.0%	0.286	[-0.073, 0.612]

	2	66.7%	0.330	[-0.093, 0.665]
	3	70.0%	0.341	[-0.023, 0.644]
<b>Trussell, Lewis, &amp; Strichter, 2008</b>		<b>87.0%</b>	<b>0.740</b>	<b>[0.580, 0.849]</b>
	1	89.5%	0.683	[0.054, 0.972]
	2	85.0%	0.688	[0.146, 0.942]
	3	79.2%	0.556	[0.064, 0.857]
	4	87.5%	0.667	[0.008, 0.970]
	5	87.0%	0.739	[0.256, 0.952]
	6	95.2%	0.883	[0.330, 1.000]
<b>Wehby, Falk, Barton- Arwood, Lane, &amp; Cooley, 2003</b>		<b>77.5%</b>	<b>0.443</b>	<b>[0.330, 0.546]</b>
	1	75.0%	0.333	[-0.246, 0.873]
	2	75.0%	0.333	[-0.246, 0.873]
	3	90.9%	0.688	[-0.121, 1.000]
	4	75.0%	0.333	[-0.246, 0.873]
	5	90.0%	0.792	[-0.034, 1.000]
	6	75.0%	0.333	[-0.246, 0.873]
	7	91.7%	0.778	[-0.007, 1.000]
	8	91.7%	0.778	[-0.007, 1.000]
	9	81.8%	0.542	[-0.181, 0.959]
	10	80.0%	0.583	[-0.213, 0.963]
	11	83.3%	0.556	[-0.146, 0.960]
	12	75.0%	0.333	[-0.246, 0.873]
	13	72.7%	0.313	[-0.284, 0.869]
	14	83.3%	0.556	[-0.146, 0.960]
	15	60.0%	0.167	[-0.490, 0.775]
	16	77.3%	0.427	[-0.082, 0.814]
	17	75.0%	0.333	[-0.200, 0.822]
	18	80.0%	0.467	[-0.077, 0.859]
	19	89.5%	0.683	[0.054, 0.972]
	20	60.0%	0.167	[-0.394, 0.683]
	21	72.7%	0.314	[-0.162, 0.741]
	22	75.0%	0.333	[-0.200, 0.822]
	23	75.0%	0.278	[-0.187, 0.762]
	24	78.9%	0.367	[-0.141, 0.831]
	25	60.0%	0.167	[-0.394, 0.683]



<b>Wilkinson, 2005</b>		<b>100.0%</b>	<b>1.000</b>	<b>[0.586, 1.000]</b>
	1	100.0%	1.000	[0.280, 1.000]
	2	100.0%	1.000	[0.280, 1.000]
<b>Willis, Whalen, Sweeney, &amp; McLaughlin, 1995</b>		<b>100.0%</b>	<b>1.000</b>	<b>[0.555, 1.000]</b>
	1	100.0%	1.000	[0.058, 1.000]
	2	100.0%	1.000	[0.266, 1.000]

*Note:* CI= 95% confidence interval; PAND= Percent of All Non-overlapping Data; Statistics in bold indicate whole study data. Statistics in numbered rows represent individual comparison analysis. Row numbers are not in any particular order.

Table 4

*Studies per Publication Year*

Year Published	Number of Studies Published ( <i>k</i> )
1986	2
1987	0
1988	1
1989	5
1990	3
1991	0
1992	5
1993	2
1994	2
1995	7
1996	3
1997	1
1998	3
1999	0
2000	1
2001	6
2002	1
2003	4
2004	4
2005	7
2006	3
2007	2
2008	6
2009	6
2010	0

*Note:* *k*=number of studies

Table 5

*Included Studies Journal Publishers Distribution*

Journal	Number of Studies Published <i>k</i>
<i>Behavioral Disorders</i>	11
<i>Journal of Emotional &amp; Behavioral Disorders</i>	9
<i>School Psychology Review</i>	9
<i>Education &amp; Treatment of Children</i>	6
<i>Behavior Modification</i>	5
<i>Child &amp; Family Behavior Therapy</i>	5
<i>Journal of Applied Behavior Analysis</i>	4
<i>Journal of Behavioral Education</i>	4
<i>Behavioral Interventions</i>	3
<i>Journal of School Psychology</i>	3
<i>Journal of Early and Intensive Behavior Intervention</i>	2
<i>Preventing School Failure</i>	2
<i>Psychology in the Schools</i>	2
<i>Behavior Therapy</i>	1
<i>Child &amp; Family Therapy</i>	1
<i>Exceptional Children</i>	1
<i>Journal of Applied Behavior Analysis</i>	1
<i>Journal of Negro Education</i>	1

<i>Journal of Positive Behavioral Interventions</i>	1
<i>Journal of Special Education Technology</i>	1
<i>Therapeutic Recreation Journal</i>	1

---

*Note:*  $k$ =number of studies

Table 6

*Effect Size Comparisons by Student Sex and Race*

<i>N</i>	Overall PAND	Overall RPhi	Confidence Interval
All (713)	87.2%	0.744 <sup>b</sup>	[0.732, 0.755]
Sex			
Male (554)	86.6%	0.731 <sup>b</sup>	[0.718, 0.744]
Female (87)	86.5%	0.731 <sup>b</sup>	[0.692, 0.766]
Sex unknown (68)	93.3%	0.865 <sup>a</sup>	[0.832, 0.892]
Race			
African American (126)	85.7%	0.711 <sup>b</sup>	[0.680, 0.740]
Hispanic (23)	92.8%	0.856 <sup>a</sup>	[0.754, 0.921]
Caucasian (135)	90.8%	0.805 <sup>a</sup>	[0.778, 0.830]
Race Unknown (429)	86.5%	0.725 <sup>b</sup>	[0.710, 0.740]

*Note:* CI= confidence interval; *N*=Number of participants in primary studies; PAND= Percent of All Non-overlapping Data; <sup>a</sup>=large effect, <sup>b</sup>=moderate effect

Table 7

*Effect Size Comparison by Student Characteristics<sup>1</sup>*

ED characteristics ( <i>N</i> )	Overall PAND	Overall RPhi	Confidence Interval
(A) Inability to Learn (424)	86.4%	0.720 <sup>b</sup>	[0.703, 0.735]
(B) Relationship Problems (97)	79.3%	0.581 <sup>b</sup>	[0.535, 0.623]
(C) Inappropriate Behavior (417)	86.7%	0.734 <sup>b</sup>	[0.718, 0.748]
(D) Unhappiness or Depression (72)	82.8%	0.628 <sup>b</sup>	[0.583, 0.670]
(E) Physical Symptoms or Fears (32)	79.4%	0.550 <sup>b</sup>	[0.477, 0.616]
Schizophrenia-Like Features (13)	91.4%	0.828 <sup>a</sup>	[0.737, 0.891]
Socially Maladjusted (39)	83.9%	0.678 <sup>b</sup>	[0.624, 0.726]

*Note:* CI= confidence interval; *N*=number of participants included in primary studies; PAND= Percent of All Non-overlapping Data; <sup>a</sup>=large effect, <sup>b</sup>=moderate effect

<sup>1</sup> Participants may have been reported as having more than one ED characteristics.

Table 8

*Effect Size Comparisons by Setting*

Setting (N)	Overall PAND	Overall RPhi	Confidence Interval
General Education (33)	95.0%	0.899 <sup>a</sup>	[0.841, 0.938]
Resource (78)	94.8%	0.889 <sup>a</sup>	[0.864, 0.910]
Self-Contained (381)	86.0%	0.720 <sup>b</sup>	[0.704, 0.736]
All Separate <sup>2</sup> (144)	88.7%	0.776 <sup>b</sup>	[0.751, 0.798]

*Note:* N=number of comparisons; PAND= Percent of All Non-overlapping Data; RPhi Robust Phi; <sup>a</sup>=large effect, <sup>b</sup>=moderate effect

<sup>1</sup> All Separate includes alternative schools, residential schools, day treatment facilities, and psychiatric facilities.

Table 9

*Effect Size Comparison by Research Design*

Research Design ( <i>k</i> )	Number of comparisons ( <i>N</i> )	Overall PAND	Overall RPhi	Confidence Interval
Multiple Baseline across Subjects (29)	300	84.4%	0.686 <sup>b</sup>	[0.667, 0.704]
Withdrawal (11)	60	84.8%	0.693 <sup>b</sup>	[0.646, 0.735]
Alternating Treatments (7)	61	88.9%	0.730 <sup>b</sup>	[0.671, 0.780]
Multiple Baseline across Probes (7)	55	97.7%	0.955 <sup>a</sup>	[0.922, 0.975]
Multiple Baseline across Settings (7)	114	91.0%	0.815 <sup>a</sup>	[0.786, 0.840]
Complex ABC (8)	40	79.1%	0.579 <sup>b</sup>	[0.519, 0.634]
Multiple Baseline across Behaviors (5)	53	90.2%	0.800 <sup>a</sup>	[0.763, 0.831]

*Note*, *k* = number of studies, *N*= number of effects, CI = confidence interval, PAND = Percent of all Non-Overlapping Data;  
<sup>a</sup> =large effect, <sup>b</sup> =moderate effect



Table 10

*Effect Size Comparison by Primary Study Quality*

(N)	Overall PAND	Overall RPhi	Confidence Interval
<b>Fidelity</b>			
Reported (281)	89.8%	0.794 <sup>b</sup>	[0.776, 0.811]
Not Reported (432)	85.7%	0.715 <sup>b</sup>	[0.699, 0.729]
<b>IOA checks in each phase</b>			
“Meets Evidence Standards”(98)	91.0%	0.811 <sup>a</sup>	[0.788, 0.833]
“Meets Evidence Standards with Reservations” (107)	92.5%	0.844 <sup>a</sup>	[0.809, 0.873]
“Does Not Meet Evidence Standards (6)	100%	0.711 <sup>b</sup>	[0.696, 0.726]
<b>&gt;20% IOA checks per comparison</b>			
No (75)	87.6%	0.750 <sup>b</sup>	[0.713, 0.783]
Yes (639)	87.1%	0.742 <sup>b</sup>	[0.730, 0.754]

*Note:* N=number of comparisons; PAND= Percent of All Non-overlapping Data; WWC=What Works Clearinghouse; <sup>a</sup>=large effect, <sup>b</sup>=moderate effect

Table 11

*Effect Size Comparisons of Interventions Used with Students with ED*

Interventions	Overall PAND	Overall RPhi	Confidence Interval
<b>Self-Management/ Monitoring (<math>k=22</math>)</b>			
Intervention only ( $n=174$ )	86.8%	0.735 <sup>b</sup>	[0.712, 0.756]
Intervention in combination ( $n=51$ )	80.1%	0.568 <sup>b</sup>	[0.538, 0.652]
Total ( $n=225$ )	86.3%	0.726 <sup>b</sup>	[0.702, 0.748]
<b>Positive reinforcement (<math>k=16</math>)</b>			
Intervention only ( $n=70$ )	88.8%	0.775 <sup>b</sup>	[0.742, 0.804]
Intervention in combination ( $n=123$ )	84.0%	0.681 <sup>b</sup>	[0.646, 0.713]
Total ( $n=193$ )	86.3%	0.725 <sup>b</sup>	[0.701, 0.747]
<b>Instruction (<math>k=11</math>)</b>			
Intervention only ( $n=90$ )	91.1%	0.896 <sup>a</sup>	[0.877, 0.913]
Intervention in combination ( $n=29$ )	93.6%	0.869 <sup>a</sup>	[0.800, 0.916]
Total ( $n=119$ )	93.5%	0.830 <sup>a</sup>	[0.804, 0.852]
<b>Direct Instruction (<math>k=7</math>)</b>			
Intervention only ( $n=43$ )	88.5%	0.756 <sup>b</sup>	[0.681, 0.817]
Intervention in combination ( $n=36$ )	88.1%	0.752 <sup>b</sup>	[0.697, 0.799]
Total ( $n=79$ )	88.2%	0.745 <sup>b</sup>	[0.711, 0.791]

Peer Tutoring ( $k=6$ )

Intervention only ( $n=32$ )	83.3%	1.000 <sup>a</sup>	[0.693, 1.000]
Intervention in combination ( $n=33$ )	68.1%	0.361 <sup>c</sup>	[0.104, 0.577]
Total ( $n=65$ )	85.7%	0.713 <sup>b</sup>	[0.672, 0.750]

Punishment ( $k=5$ )

Intervention only ( $n=0$ )	NA	NA	NA
Intervention in combination ( $n=36$ )	87.6%	0.736 <sup>b</sup>	[0.633, 0.796]
Total ( $n=0$ )	NA	NA	NA

Function Based Interventions ( $k=4$ )

Intervention only ( $n=14$ )	87.9%	0.744 <sup>b</sup>	[0.624, 0.832]
Intervention in combination ( $n=7$ )	70.8%	0.417 <sup>a</sup>	[-0.066, 0.760]
Total ( $n=21$ )	82.0%	0.640 <sup>b</sup>	[0.531, 0.730]

Social Skill Instruction ( $k=4$ )

Intervention only ( $n=35$ )	85.0%	0.681 <sup>b</sup>	[0.615, 0.738]
Intervention in combination ( $n=0$ )	NA	NA	NA
Total ( $n=35$ )	85.0%	0.681 <sup>b</sup>	[0.615, 0.738]

Peer Modeling ( $k=3$ )

Intervention only ( $n=19$ )	92.4%	0.848 <sup>a</sup>	[0.768, 0.904]
Intervention in combination ( $n=0$ )	NA	NA	NA
Total ( $n=19$ )	92.4%	0.848 <sup>a</sup>	[0.768, 0.904]

Curricular Modifications ( $k=2$ )

Intervention only ( $n=30$ )	79.7%	0.591 <sup>b</sup>	[0.514, 0.659]
Intervention in combination ( $n=0$ )	NA	NA	NA
Total ( $n=30$ )	79.7%	0.591 <sup>b</sup>	[0.514, 0.659]
Computer Assisted Instruction ( $k=1$ )			
Intervention only ( $n=27$ )	98.7%	0.756 <sup>b</sup>	[0.681, 0.817]
Intervention in combination ( $n=6$ )	100%	1.000 <sup>a</sup>	[0.693, 1.000]
Total ( $n=33$ )	98.8%	1.000 <sup>a</sup>	[0.693, 1.000]
Choice ( $k=1$ )			
Intervention only ( $n=15$ )	84.4%	0.469 <sup>c</sup>	[0.366, 0.611]
Intervention in combination ( $n=0$ )	NA	NA	NA
Total ( $n=15$ )	84.4%	0.469 <sup>c</sup>	[0.366, 0.611]

---

*Note:*  $k$ =number of studies;  $n$ =number of comparisons; PAND= Percent of All Non overlapping Data; RPhi =Robust Phi; <sup>a</sup>=large effect, <sup>b</sup>=moderate effect, <sup>c</sup>=small effect

Table 12

*Effect Size Comparisons of Interventions Implemented with Students with the Inability to Learn Characteristic<sup>1</sup>*

Interventions (n)	Overall PAND	Overall RPhi	Confidence Interval
<b>Self-Management/ Monitoring</b>			
Intervention only (115)	90.0%	0.798 <sup>b</sup>	[0.771, 0.822]
Intervention in combination (38)	75.7%	0.505 <sup>b</sup>	[0.434, 0.570]
Total (153)	86.8%	0.732 <sup>b</sup>	[0.707, 0.756]
<b>Positive reinforcement</b>			
Intervention only (18)	68.1%	0.303 <sup>c</sup>	[0.171, 0.427]
Intervention in combination (66)	80.0%	0.592 <sup>b</sup>	[0.539, 0.641]
Total (84)	77.7%	0.541 <sup>b</sup>	[0.492, 0.587]
<b>Instruction</b>			
Intervention only (76)	90.3%	0.798 <sup>b</sup>	[0.764, 0.827]
Intervention in combination (20)	96.1%	0.917 <sup>a</sup>	[0.835, 0.962]
Total (96)	91.0%	0.816 <sup>a</sup>	[0.787, 0.842]
<b>Direct Instruction</b>			
Intervention only (29)	88.6%	0.758 <sup>b</sup>	[0.667, 0.828]
Intervention in combination (10)	79.7%	0.514 <sup>b</sup>	[0.328, 0.667]

Total (39)	85.9%	0.692 <sup>a</sup>	[0.611, 0.759]
<b>Peer Tutoring</b>			
Intervention only (22)	84.2%	0.667 <sup>b</sup>	[0.585, 0.736]
Intervention in combination (6)	75.9%	0.371 <sup>c</sup>	[0.156, 0.563]
Total (28)	82.5%	0.648 <sup>b</sup>	[0.577, 0.709]
<b>Punishment</b>			
Intervention only (0)	NA	NA	NA
Intervention in combination (0)	NA	NA	NA
Total (0)	NA	NA	NA
<b>Function Based Interventions</b>			
Intervention only (12)	85.4%	0.708 <sup>b</sup>	[0.573, 0.808]
Intervention in combination (5)	85.4%	0.706 <sup>b</sup>	[0.405, 0.882]
Total (17)	85.4%	0.708 <sup>b</sup>	[0.592, 0.797]
<b>Social Skill Instruction</b>			
Intervention only (6)	87.3%	0.702 <sup>b</sup>	[0.542, 0.818]
Intervention in combination (0)	NA	NA	NA
Total (6)	87.3%	0.702 <sup>b</sup>	[0.542, 0.818]
<b>Peer Modeling</b>			
Intervention only (0)	NA	NA	NA
Intervention in combination (0)	NA	NA	NA
Total (0)	NA	NA	NA

## Curricular Modifications

Intervention only (30)	79.7%	0.591 <sup>b</sup>	[0.514, 0.659]
Intervention in combination (0)	NA	NA	NA
Total (30)	79.7%	0.591 <sup>b</sup>	[0.514, 0.659]

## Computer Assisted Instruction

Intervention only (6)	97.0%	0.941 <sup>a</sup>	[0.618, 1.000]
Intervention in combination (4)	100%	1.000 <sup>a</sup>	[0.564, 1.000]
Total (10)	98.2%	0.964 <sup>a</sup>	[0.758, 1.000]

## Choice

Intervention only (15)	84.4%	0.496 <sup>c</sup>	[0.366, 0.611]
Intervention in combination (0)	NA	NA	NA
Total (15)	84.4%	0.496 <sup>c</sup>	[0.366, 0.611]

---

*Note:* CI= confidence interval; *n*=number of comparisons; PAND= Percent of All Non-overlapping Data; <sup>a</sup>=large effect, <sup>b</sup>=moderate effect, <sup>c</sup>=small effect

<sup>1</sup> Participants may have also been reported as having other ED characteristics in addition to Inability to Learn.

Table 13

*Effect Size Comparisons of Interventions Implemented with Students with Relationship Problems<sup>1</sup>*

Interventions ( <i>n</i> )	Overall PAND	Overall RPhi	Confidence Interval
<b>Self-Management/Monitoring</b>			
Intervention only (11)	89.1%	0.769 <sup>b</sup>	[0.644, 0.857]
Intervention in combination (9)	75.5%	0.507 <sup>b</sup>	[0.353, 0.636]
Total (20)	82.6%	0.644 <sup>b</sup>	[0.549, 0.723]
<b>Positive reinforcement</b>			
Intervention only (16)	65.5%	0.261 <sup>c</sup>	[0.122, 0.392]
Intervention in combination (38)	80.2%	0.577 <sup>b</sup>	[0.496, 0.650]
Total (54)	75.6%	0.508 <sup>b</sup>	[0.440, 0.570]
<b>Instruction</b>			
Intervention only (0)	NA	NA	NA
Intervention in combination (0)	NA	NA	NA
Total (0)	NA	NA	NA
<b>Direct Instruction</b>			
Intervention only (6)	78.5%	0.529 <sup>b</sup>	[0.253, 0.735]
Intervention in combination (4) <sup>2</sup>	73.5%	0.363 <sup>c</sup>	[0.083, 0.605]



## Direct Instruction

Intervention only (6)	78.5%	0.529 <sup>b</sup>	[0.253, 0.735]
Intervention in combination (4) <sup>2</sup>	73.5%	0.363 <sup>c</sup>	[0.083, 0.605]
Total (10)	75.9%	0.450 <sup>c</sup>	[0.261, 0.611]

## Peer Tutoring

Intervention only (6)	100%	1.000 <sup>a</sup>	[0.851, 1.000]
Intervention in combination (4) <sup>2</sup>	73.5%	0.363 <sup>c</sup>	[0.083, 0.605]
Total (10)	87.4%	0.746 <sup>b</sup>	[0.600, 0.848]

## Punishment

Intervention only (0)	NA	NA	NA
Intervention in combination (28)	84.3%	0.639 <sup>b</sup>	[0.534, 0.726]
Total (28)	84.3%	0.639 <sup>b</sup>	[0.534, 0.726]

## Function Based Interventions

Intervention only (0)	NA	NA	NA
Intervention in combination (0)	NA	NA	NA
Total (0)	NA	NA	NA

## Social Skill Instruction

Intervention only (13)	78.9%	0.566 <sup>b</sup>	[0.434, 0.676]
Intervention in combination (0)	NA	NA	NA
Total (13)	78.9%	0.566 <sup>b</sup>	[0.434, 0.676]

## Peer Modeling

## Curricular Modifications

Intervention only (2)	100%	1.000 <sup>a</sup>	[0.620, 1.000]
Intervention in combination (0)	NA	NA	NA
Total (2)	100%	1.000 <sup>a</sup>	[0.620, 1.000]

## Computer Assisted Instruction

Intervention only (0)	NA	NA	NA
Intervention in combination (0)	NA	NA	NA
Total (0)	NA	NA	NA

## Choice

Intervention only (0)	NA	NA	NA
Intervention in combination (0)	NA	NA	NA
Total (0)	NA	NA	NA

---

*Note:* CI= confidence interval; *n*=number of comparisons; PAND= Percent of All Non-overlapping Data; <sup>a</sup>=large effect, <sup>b</sup>=moderate effect, <sup>c</sup>=small effect

<sup>1</sup> Participants may have also been reported as having other ED characteristics in addition to Relationship Problems.

<sup>2</sup> same subset of comparisons

Table 14

*Effect Size Comparison of Interventions Implemented with Students with Inappropriate Behavior<sup>1</sup>*

Interventions ( <i>n</i> )	Overall PAND	Overall RPhi	Confidence Interval
<b>Self-Management/Monitoring</b>			
Intervention only (109)	85.1%	0.702 <sup>b</sup>	[0.670 - 0.731]
Intervention in combination (45)	78.4%	0.561 <sup>b</sup>	[0.498 - 0.617]
Total (154)	83.3%	0.665 <sup>b</sup>	[0.637 - 0.691]
<b>Positive reinforcement</b>			
Intervention only (44)	90.3%	0.840 <sup>a</sup>	[0.756, 0.843]
Intervention in combination (70)	81.2%	0.625 <sup>b</sup>	[0.573, 0.672]
Total (114)	85.3%	0.705 <sup>b</sup>	[0.670, 0.737]
<b>Instruction</b>			
Intervention only (31)	93.8%	0.844 <sup>a</sup>	[0.797, 0.881]
Intervention in combination (11)	91.9%	0.838 <sup>a</sup>	[0.728, 0.909]
Total (42)	93.5%	0.850 <sup>a</sup>	[0.811, 0.882]
<b>Direct Instruction</b>			
Intervention only (28)	88.1%	0.743 <sup>b</sup>	[0.652, 0.814]
Intervention in combination (23)	88.4%	0.744 <sup>b</sup>	[0.674 - 0.801]

## Direct Instruction

Intervention only (28)	88.1%	0.743 <sup>b</sup>	[0.652, 0.814]
Intervention in combination (23)	88.4%	0.744 <sup>b</sup>	[0.674 - 0.801]
Total (51)	88.2%	0.743 <sup>b</sup>	[0.691, 0.789]

## Peer Tutoring

Intervention only (12)	84.5%	0.676 <sup>b</sup>	[0.551, 0.773]
Intervention in combination (23)	88.4%	0.744 <sup>b</sup>	[0.674, 0.801]
Total (35)	87.3%	0.739 <sup>b</sup>	[0.683, 0.786]

## Punishment

Intervention only (0)	NA	NA	NA
Intervention in combination (24)	94.0%	0.826 <sup>a</sup>	[0.698, 0.907]
Total (24)	94.0%	0.826 <sup>a</sup>	[0.698, 0.907]

## Function Based Interventions

Intervention only (6)	87.0%	0.740 <sup>b</sup>	[0.580, 0.849]
Intervention in combination (2)	100%	1.000 <sup>a</sup>	[0.586, 1.000]
Total (8)	89.1%	0.782 <sup>b</sup>	[0.645, 0.874]

## Social Skill Instruction

Intervention only (30)	86.4%	0.717 <sup>b</sup>	[0.653, 0.771]
Intervention in combination (0)	NA	NA	NA
Total (30)	86.4%	0.717 <sup>b</sup>	[0.653, 0.771]

## Peer Modeling

## Curricular Modifications

Intervention only (12)	80.8%	0.614 <sup>b</sup>	[0.500, 0.707]
Intervention in combination (0)	NA	NA	NA
Total (12)	80.8%	0.614 <sup>b</sup>	[0.500, 0.707]

## Computer Assisted Instruction

Intervention only (0)	NA	NA	NA
Intervention in combination (0)	NA	NA	NA
Total (0)	NA	NA	NA

## Choice

Intervention only (15)	84.4%	0.502 <sup>b</sup>	[0.373, 0.616]
Intervention in combination (0)	NA	NA	NA
Total (15)	84.4%	0.502 <sup>b</sup>	[0.373, 0.616]

---

*Note:* CI= confidence interval; *n*=number of comparisons; PAND= Percent of All Non-overlapping Data; <sup>a</sup>=large effect, <sup>b</sup>=moderate effect

<sup>1</sup> Participants may have also been reported as having other ED characteristics in addition to Inappropriate Behavior.

Table 15

*Effect Size Comparisons of Interventions Implemented with Students with Unhappiness or Depression<sup>1</sup>*

Interventions ( <i>n</i> )	Overall PAND	Overall RPhi	Confidence Interval
<b>Self-Management/Monitoring</b>			
Intervention only (10)	71.9%	0.419 <sup>c</sup>	[0.316, 0.514]
Intervention in combination (3)	100%	1.00 <sup>a</sup>	[0.789, 1.000]
Total (13)	75.5%	0.491 <sup>c</sup>	[0.398, 0.575]
<b>Positive reinforcement</b>			
Intervention only (4)	88.9%	0.777 <sup>b</sup>	[0.587, 0.892]
Intervention in combination (11)	88.0%	0.757 <sup>b</sup>	[0.625, 0.849]
Total (15)	88.3%	0.766 <sup>b</sup>	[0.677, 0.840]
<b>Instruction</b>			
Intervention only (5)	93.8%	0.867 <sup>a</sup>	[0.696, 0.952]
Intervention in combination (5)	96.3%	0.917 <sup>a</sup>	[0.664, 0.993]
Total (10)	93.8%	0.867 <sup>a</sup>	[0.696, 0.952]
<b>Direct Instruction</b>			
Intervention only (6)	78.9%	0.442 <sup>c</sup>	[0.158, 0.675]

Intervention in combination (9)	88.9%	0.709 <sup>b</sup>	[0.566, 0.815]
Total (15)	86.3%	0.639 <sup>b</sup>	[0.514, 0.741]
<b>Peer Tutoring</b>			
Intervention only (4)	77.4%	0.538 <sup>b</sup>	[0.309, 0.714]
Intervention in combination (12)	89.2%	0.717 <sup>b</sup>	[0.584, 0.816]
Total (16)	86.0%	0.668 <sup>b</sup>	[0.560, 0.755]
<b>Punishment</b>			
Intervention only (0)	NA	NA	NA
Intervention in combination (0)	NA	NA	NA
Total (0)	NA	NA	NA
<b>Function Based Interventions</b>			
Intervention only (0)	NA	NA	NA
Intervention in combination (0)	NA	NA	NA
Total (0)	NA	NA	NA
<b>Social Skill Instruction</b>			
Intervention only (15)	82.2%	0.644 <sup>b</sup>	[0.518, 0.745]
Intervention in combination (0)	NA	NA	NA
Total (15)	82.2%	0.644 <sup>b</sup>	[0.518, 0.745]
<b>Peer Modeling</b>			
Intervention only (0)	NA	NA	NA
Intervention in combination (0)	NA	NA	NA

Total (0)	NA	NA	NA
Curricular Modifications			
Intervention only (3)	94.7%	0.864 <sup>a</sup>	[0.581, 0.975]
Intervention in combination (0)	NA	NA	NA
Total (3)	94.7%	0.864 <sup>a</sup>	[0.581, 0.975]
Computer Assisted Instruction			
Intervention only (0)	NA	NA	NA
Intervention in combination (0)	NA	NA	NA
Total (0)	NA	NA	NA
Choice			
Intervention only (10)	82.2%	0.486 <sup>c</sup>	[0.330, 0.621]
Intervention in combination (0)	NA	NA	NA
Total (10)	82.2%	0.486 <sup>c</sup>	[0.330, 0.621]

---

*Note:* CI= confidence interval; *n*=number of comparisons; PAND= Percent of All Non-overlapping Data; <sup>a</sup>=large effect, <sup>b</sup>=moderate effect, <sup>c</sup>=small effect

<sup>l</sup> Participants may have also been reported as having other ED characteristics in addition to Unhappiness or Depression.



Table 16

*Effect Size Comparisons of Interventions Implemented with Students with Physical Symptoms or Fears<sup>1</sup>*

Interventions ( <i>n</i> )	Overall PAND	Overall RPhi	Confidence Interval
<b>Self-Management/Monitoring</b>			
Intervention only (10)	74.9%	0.476 <sup>c</sup>	[0.359, 0.580]
Intervention in combination (0)	NA	NA	NA
Total (10)	74.9%	0.476 <sup>c</sup>	[0.359, 0.580]
<b>Positive reinforcement</b>			
Intervention only (2)	100%	1.00 <sup>a</sup>	[0.606, 1.000]
Intervention in combination (6)	88%	0.760 <sup>b</sup>	[0.543, 0.890]
Total (8)	91.7%	0.832 <sup>a</sup>	[0.673, 0.923]
<b>Instruction</b>			
Intervention only (3)	80%	0.524 <sup>b</sup>	[-0.028, 0.875]
Intervention in combination (3)	91.7%	0.778 <sup>b</sup>	[0.240, 0.980]
Total (6)	86.4%	0.656 <sup>b</sup>	[0.292, 0.874]
<b>Direct Instruction</b>			
Intervention only (3)	77.8%	0.407 <sup>c</sup>	[0.005, 0.736]
Intervention in combination (2)	75.0%	0.370 <sup>c</sup>	[0.015, 0.672]

Total (5)	76.3%	0.387 <sup>c</sup>	[0.125, 0.614]
Peer Tutoring			
Intervention only (0)	NA	NA	NA
Intervention in combination (6)	80.9%	0.509 <sup>b</sup>	[0.221, 0.730]
Total (6)	80.9%	0.509 <sup>b</sup>	[0.221, 0.730]
Punishment			
Intervention only (0)	NA	NA	NA
Intervention in combination (0)	NA	NA	NA
Total (0)	NA	NA	NA
Function Based Interventions			
Intervention only (0)	NA	NA	NA
Intervention in combination (0)	NA	NA	NA
Total (0)	NA	NA	NA
Social Skill Instruction			
Intervention only (0)	NA	NA	NA
Intervention in combination (0)	NA	NA	NA
Total (0)	NA	NA	NA
Peer Modeling			
Intervention only (0)	NA	NA	NA
Intervention in combination (0)	NA	NA	NA
Total (0)	NA	NA	NA

## Curricular Modifications

Intervention only (0)	NA	NA	NA
Intervention in combination (0)	NA	NA	NA
Total (0)	NA	NA	NA

## Computer Assisted Instruction

Intervention only (0)	NA	NA	NA
Intervention in combination (0)	NA	NA	NA
Total (0)	NA	NA	NA

## Choice

Intervention only (5)	88.3%	0.580 <sup>b</sup>	[0.337, 0.764]
Intervention in combination (0)	NA	NA	NA
Total (5)	88.3%	0.580 <sup>b</sup>	[0.337, 0.764]

---

*Note:* CI= confidence interval; *n*=number of comparisons; PAND= Percent of All Non-overlapping Data; <sup>a</sup> =large effect, <sup>b</sup> =moderate effect, <sup>c</sup> =small effect

<sup>1</sup> Participants may have also been reported as having other ED characteristics in addition to Physical Symptoms or Fears.

Table 17

*Effect Size Comparisons of Interventions Implemented with Students having  
Schizophrenia-Like Features<sup>1</sup>*

Interventions ( <i>n</i> )	Overall PAND	Overall RPhi	Confidence Interval
<b>Self-Management/Monitoring</b>			
Intervention only (0)	NA	NA	NA
Intervention in combination (2)	100%	1.00 <sup>c</sup>	[0.693, 1.000]
Total (2)	100%	1.00 <sup>c</sup>	[0.693, 1.000]
<b>Positive reinforcement</b>			
Intervention only (0)	NA	NA	NA
Intervention in combination (5)	100%	1.00 <sup>c</sup>	[0.829, 1.000]
Total (5)	100%	1.00 <sup>c</sup>	[0.829, 1.000]
<b>Instruction</b>			
Intervention only (2)	100%	1.000 <sup>c</sup>	[0.655, 1.000]
Intervention in combination (2)	100%	1.000 <sup>c</sup>	[0.655, 1.000]
Total (4)	100%	1.000 <sup>c</sup>	[0.796, 1.000]
<b>Direct Instruction</b>			
Intervention only (0)	NA	NA	NA
Intervention in combination (5) <sup>2</sup>	88.3%	0.751 <sup>b</sup>	[0.603, 0.853]

Total (5)	88.3%	0.751 <sup>b</sup>	[0.603, 0.853]
Peer Tutoring			
Intervention only (0)	NA	NA	NA
Intervention in combination (5) <sup>2</sup>	88.3%	0.751 <sup>b</sup>	[0.603, 0.853]
Total (5)	88.3%	0.751 <sup>b</sup>	[0.603, 0.853]
Punishment			
Intervention only (0)	NA	NA	NA
Intervention in combination (0)	NA	NA	NA
Total (0)	NA	NA	NA
Function Based Interventions			
Intervention only (0)	NA	NA	NA
Intervention in combination (0)	NA	NA	NA
Total (0)	NA	NA	NA
Social Skill Instruction			
Intervention only (2)	77.3%	0.542 <sup>b</sup>	[0.032, 0.853]
Intervention in combination (0)	NA	NA	NA
Total (2)	77.3%	0.542 <sup>b</sup>	[0.032, 0.853]
Peer Modeling			
Intervention only (0)	NA	NA	NA
Intervention in combination (0)	NA	NA	NA
Total (0)	NA	NA	NA

## Curricular Modifications

Intervention only (0)	NA	NA	NA
Intervention in combination (0)	NA	NA	NA
Total (0)	NA	NA	NA

## Computer Assisted Instruction

Intervention only (0)	NA	NA	NA
Intervention in combination (0)	NA	NA	NA
Total (0)	NA	NA	NA

## Choice

Intervention only (0)	NA	NA	NA
Intervention in combination (0)	NA	NA	NA
Total (0)	NA	NA	NA

---

*Note:* CI= confidence interval; *n*=number of comparisons; PAND= Percent of All Non-overlapping Data; <sup>b</sup>=moderate effect, <sup>c</sup>=small effect

<sup>1</sup> Participants may have also been reported as having other ED characteristics in addition to Schizophrenia-Like Features.

<sup>2</sup> Same subset of studies.

Table 18

*Effect Size Comparisons of Interventions Implemented with Students characterized as Socially Maladjusted<sup>1</sup>*

Interventions ( <i>n</i> )	Overall PAND	Overall RPhi	Confidence Interval
<b>Self-Management/Monitoring</b>			
Intervention only (10)	70.8%	0.411 <sup>c</sup>	[0.284, 0.525]
Intervention in combination (0)	NA	NA	NA
Total (10)	70.8%	0.411 <sup>c</sup>	[0.284, 0.525]
<b>Positive reinforcement</b>			
Intervention only (4)	100%	1.000 <sup>a</sup>	[0.902, 1.000]
Intervention in combination (0)	NA	NA	NA
Total (4)	100%	1.000 <sup>a</sup>	[0.902, 1.000]
<b>Instruction</b>			
Intervention only (0)	NA	NA	NA
Intervention in combination (0)	NA	NA	NA
Total (0)	NA	NA	NA
<b>Direct Instruction</b>			
Intervention only (0)	NA	NA	NA
Intervention in combination (10)	91.5%	0.826 <sup>a</sup>	[0.740, 0.888]

Total (10)	91.5%	0.826 <sup>a</sup>	[0.740, 0.888]
<b>Peer Tutoring</b>			
Intervention only (10)	85.4%	0.707 <sup>b</sup>	[0.569, 0.809]
Intervention in combination (10)	91.5%	0.826 <sup>a</sup>	[0.740, 0.888]
Total (20)	89.2%	0.783 <sup>b</sup>	[0.714, 0.839]
<b>Punishment</b>			
Intervention only (0)	NA	NA	NA
Intervention in combination (0)	NA	NA	NA
Total (0)	NA	NA	NA
<b>Function Based Interventions</b>			
Intervention only (0)	NA	NA	NA
Intervention in combination (0)	NA	NA	NA
Total (0)	NA	NA	NA
<b>Social Skill Instruction</b>			
Intervention only (5)	68.8%	0.369 <sup>c</sup>	[0.094, 0.598]
Intervention in combination (0)	NA	NA	NA
Total (5)	68.8%	0.369 <sup>c</sup>	[0.094, 0.598]
<b>Peer Modeling</b>			
Intervention only (0)	NA	NA	NA
Intervention in combination (0)	NA	NA	NA
Total (0)	NA	NA	NA



## Curricular Modifications

Intervention only (0)	NA	NA	NA
Intervention in combination (0)	NA	NA	NA
Total (0)	NA	NA	NA

## Computer Assisted Instruction

Intervention only (0)	NA	NA	NA
Intervention in combination (0)	NA	NA	NA
Total (0)	NA	NA	NA

## Choice

Intervention only (0)	NA	NA	NA
Intervention in combination (0)	NA	NA	NA
Total (0)	NA	NA	NA

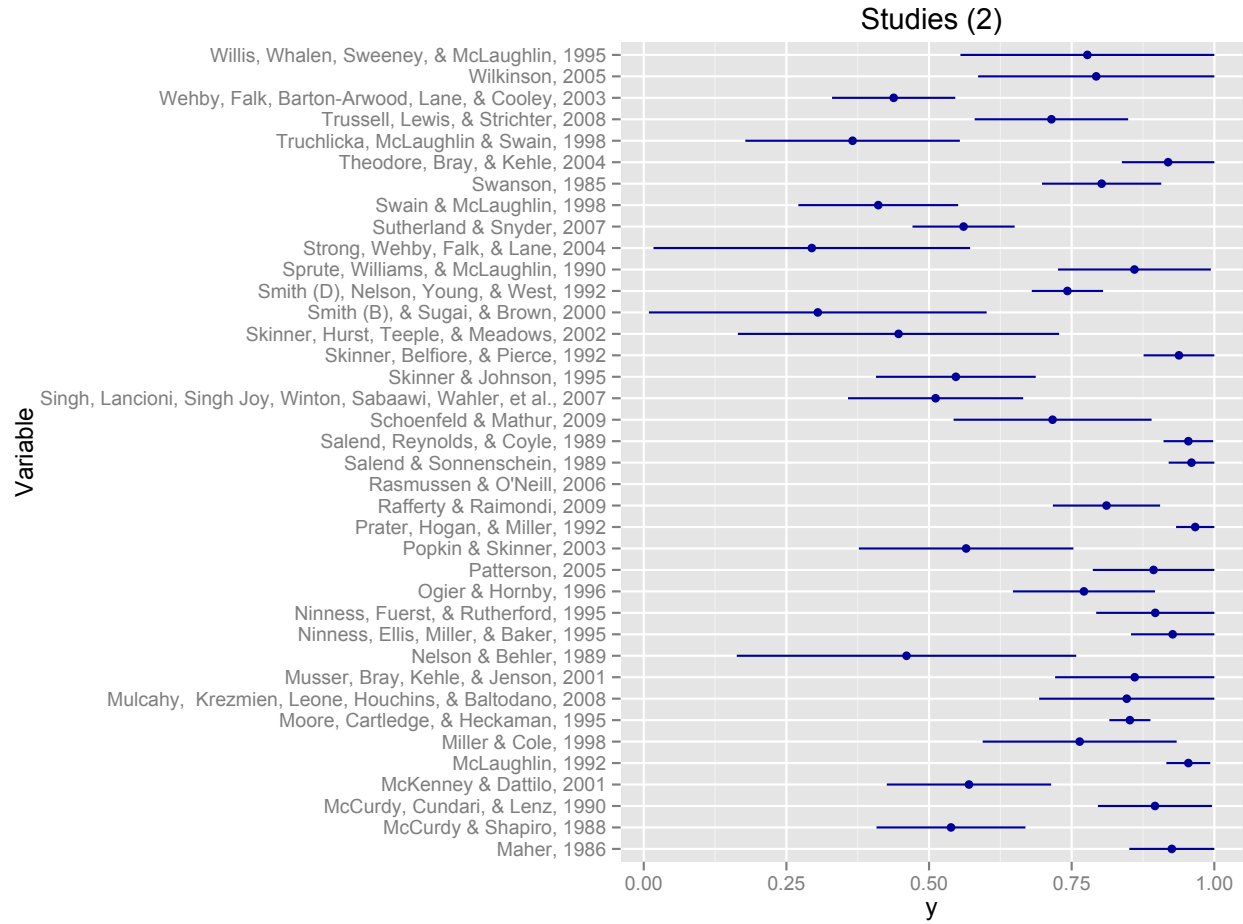
---

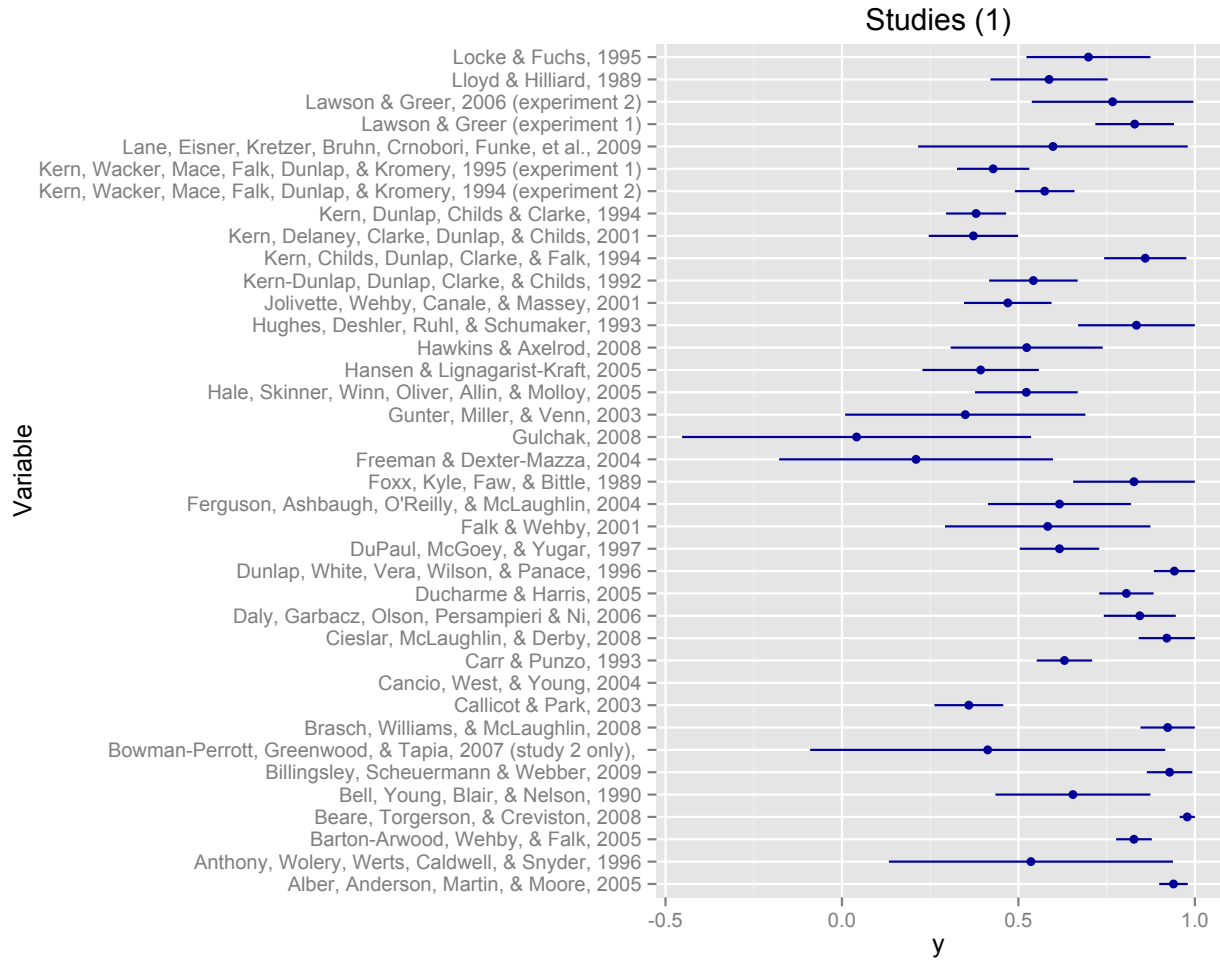
*Note:* CI= confidence interval; *n*=number of comparisons; PAND= Percent of All Non-overlapping Data; <sup>a</sup> =large effect, <sup>b</sup> =moderate effect, <sup>c</sup> =small effect

<sup>1</sup> Participants may have also been reported as having other ED characteristics in addition to being Socially Maladjusted.

**Appendix E****Figures**

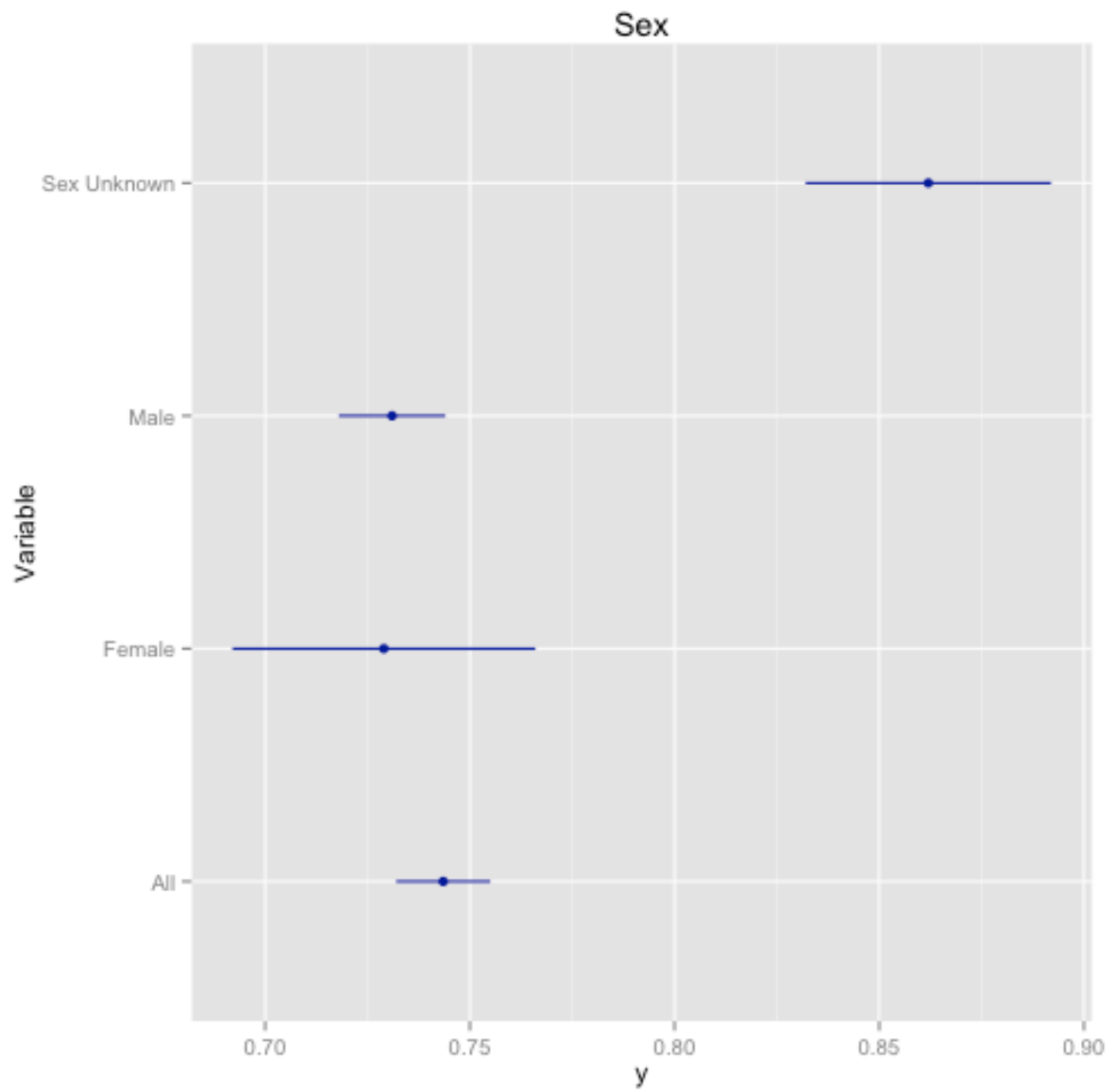
Figure 1 Individual Study and Comparison Effective Sizes





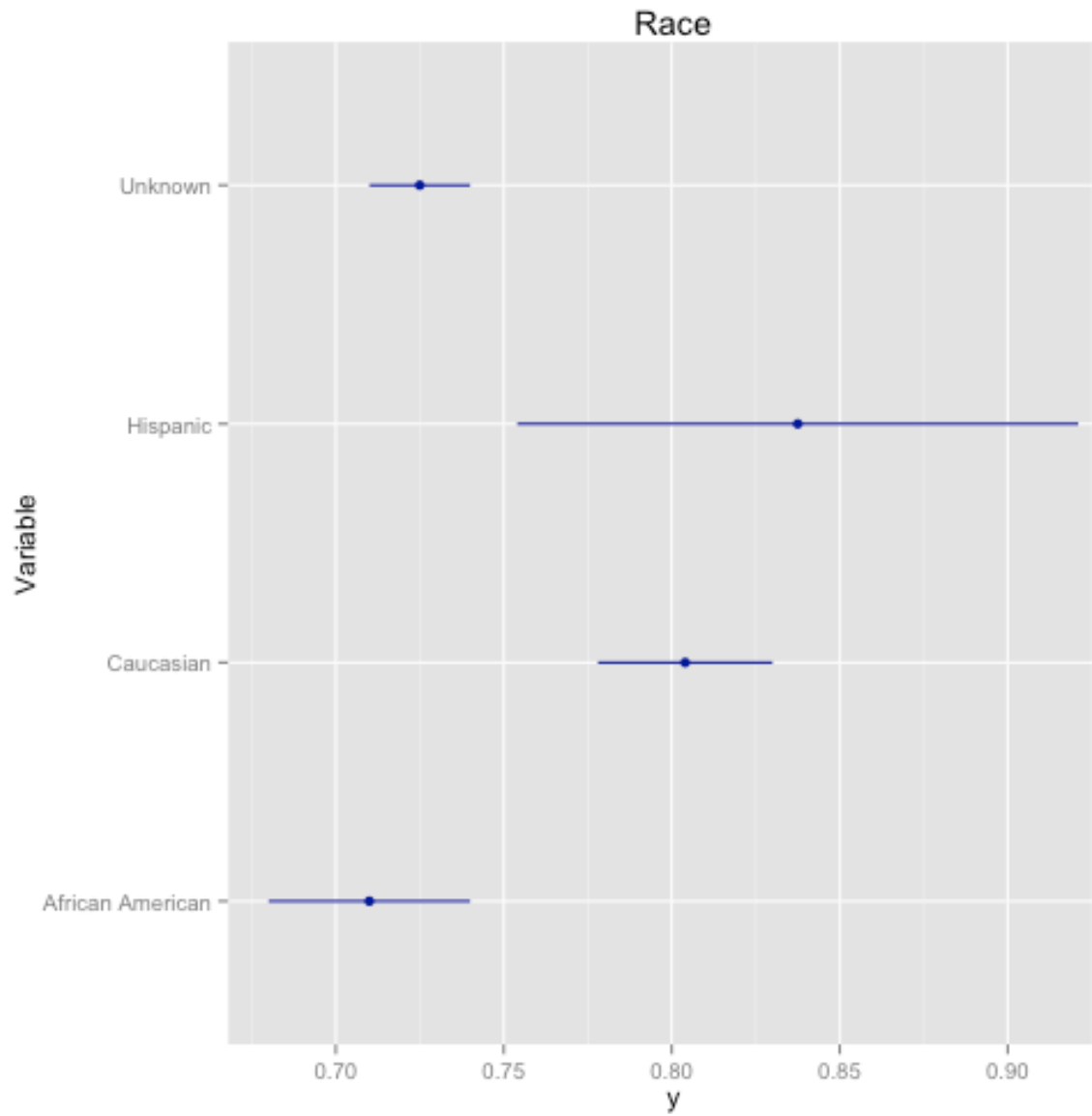
Note: RPhi with CI= 95% confidence interval.

Figure 2 Effect Size Comparisons by Student Sex

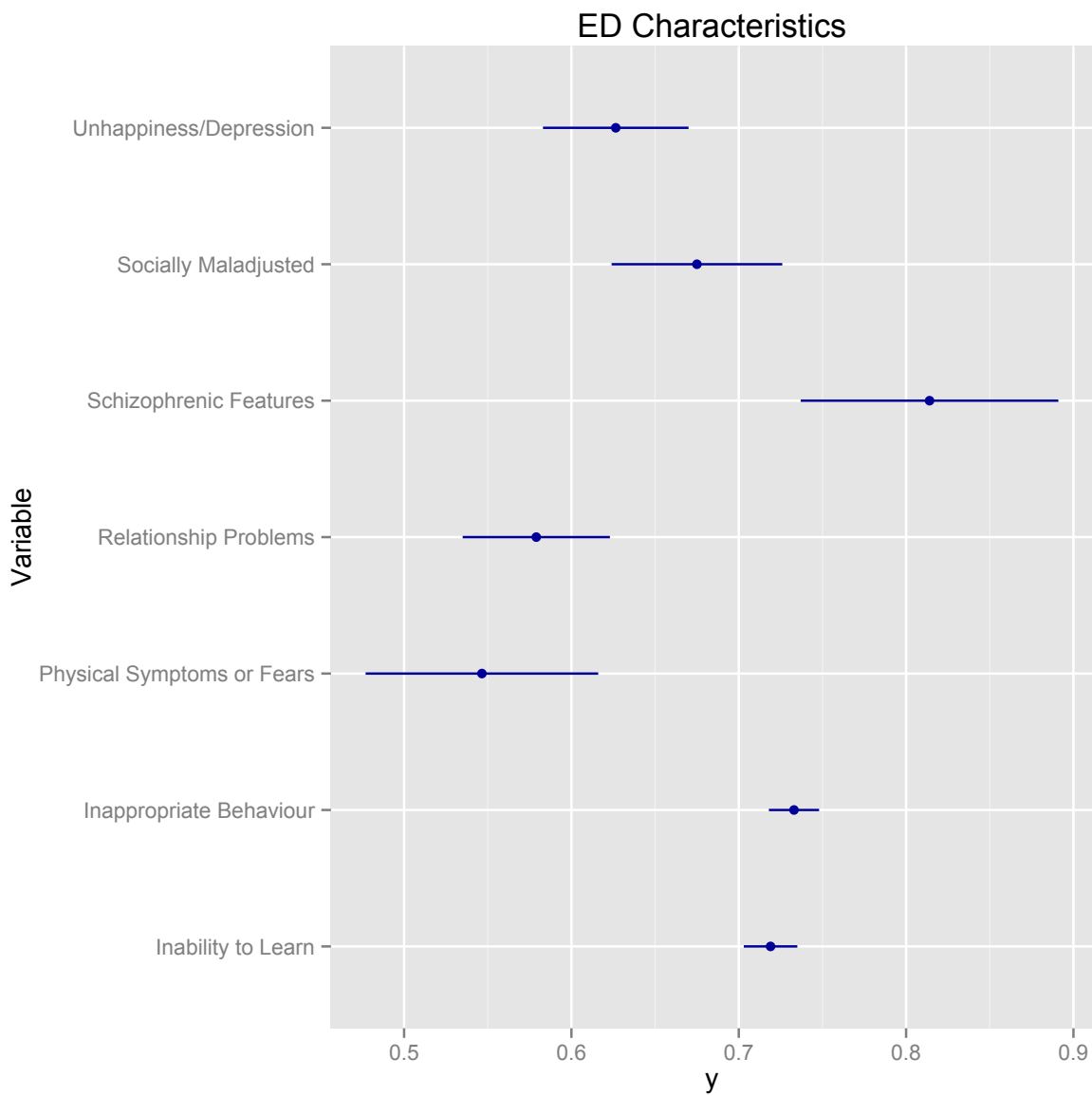


Note: RPhi with CI= 95% confidence interval.

Figure 3 Effect Size Comparisons by Student Race



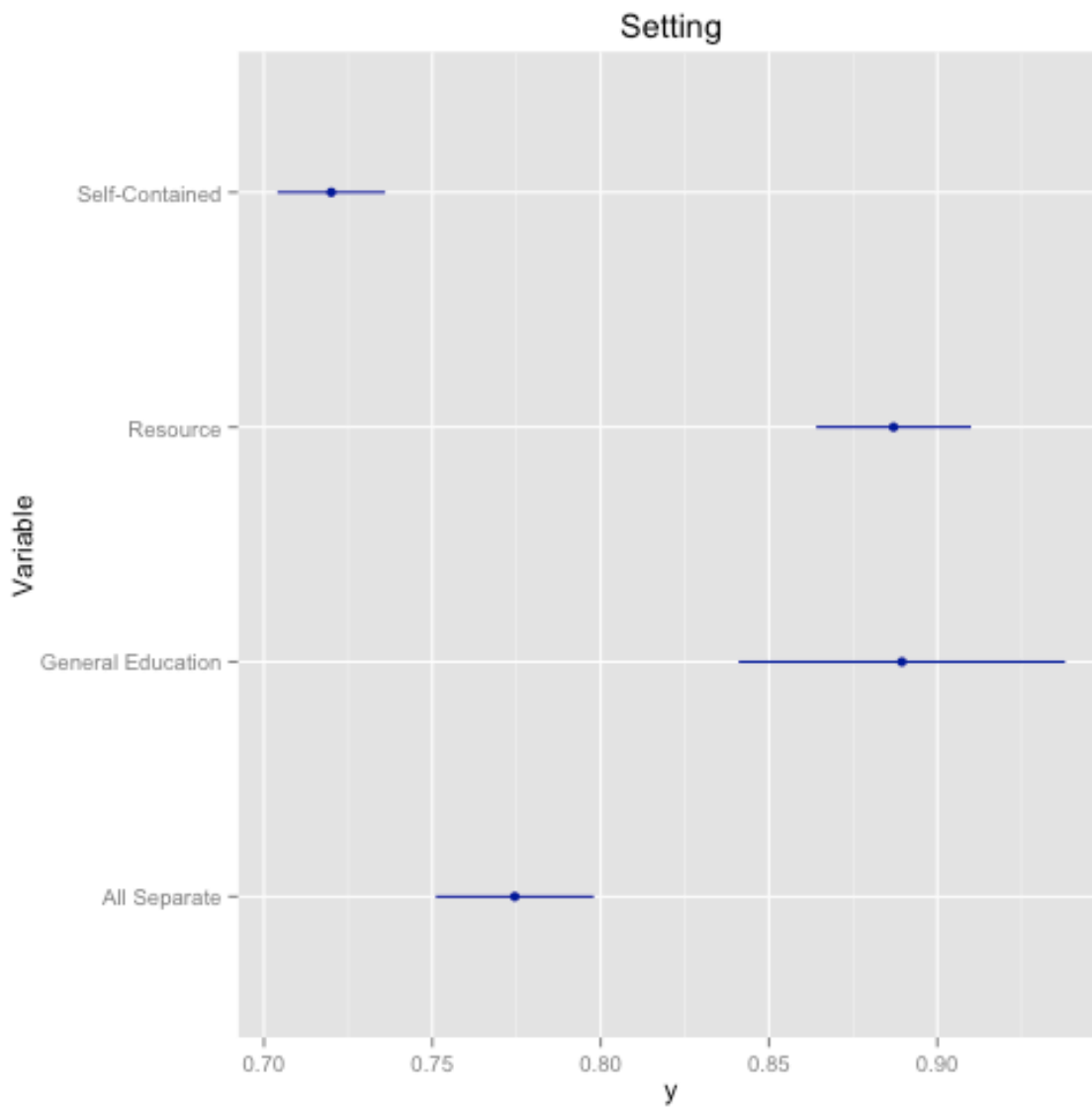
*Note:* RPhi with CI= 95% confidence interval.

Figure 4 Effect Size Comparison by Student Characteristics<sup>1</sup>

*Note:* RPhi with CI= 95% confidence interval.

<sup>1</sup> Participants may have been reported as having more than one ED characteristics.

Figure 5 Effect Size Comparisons by Setting

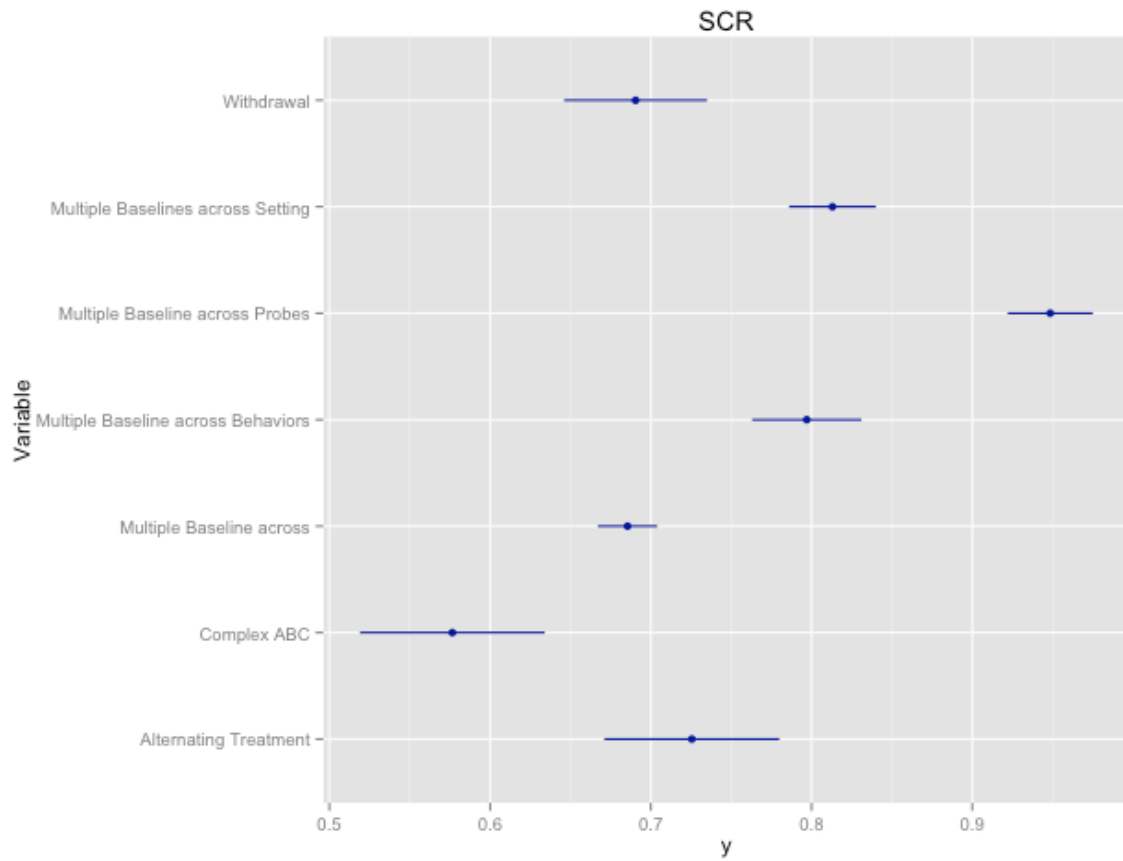


*Note:* RPhi with CI= 95% confidence interval.

<sup>1</sup> All Separate includes alternative schools, residential schools, day treatment facilities, and psychiatric facilities.

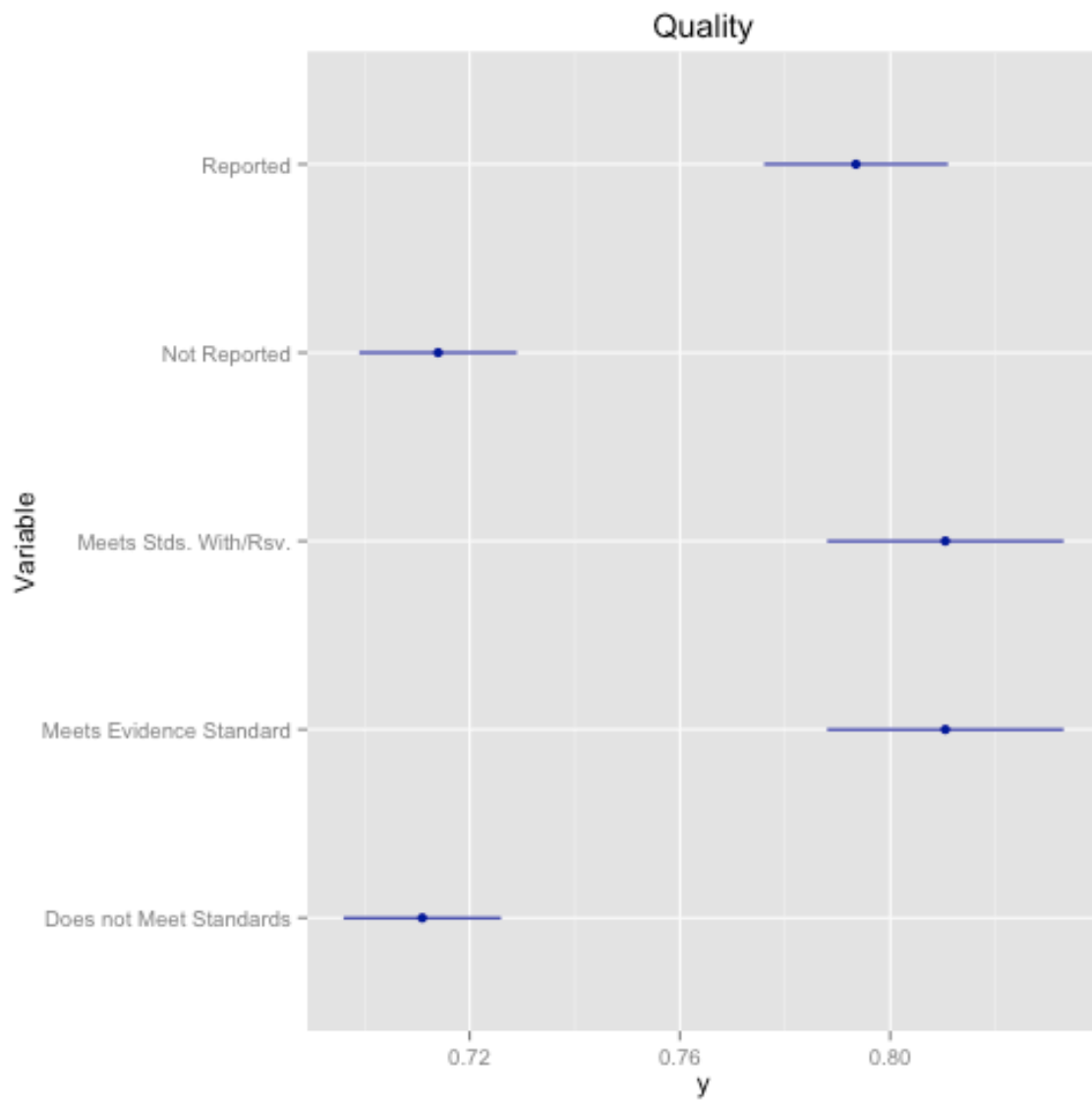


Figure 6 Effect Size Comparison by Research Design



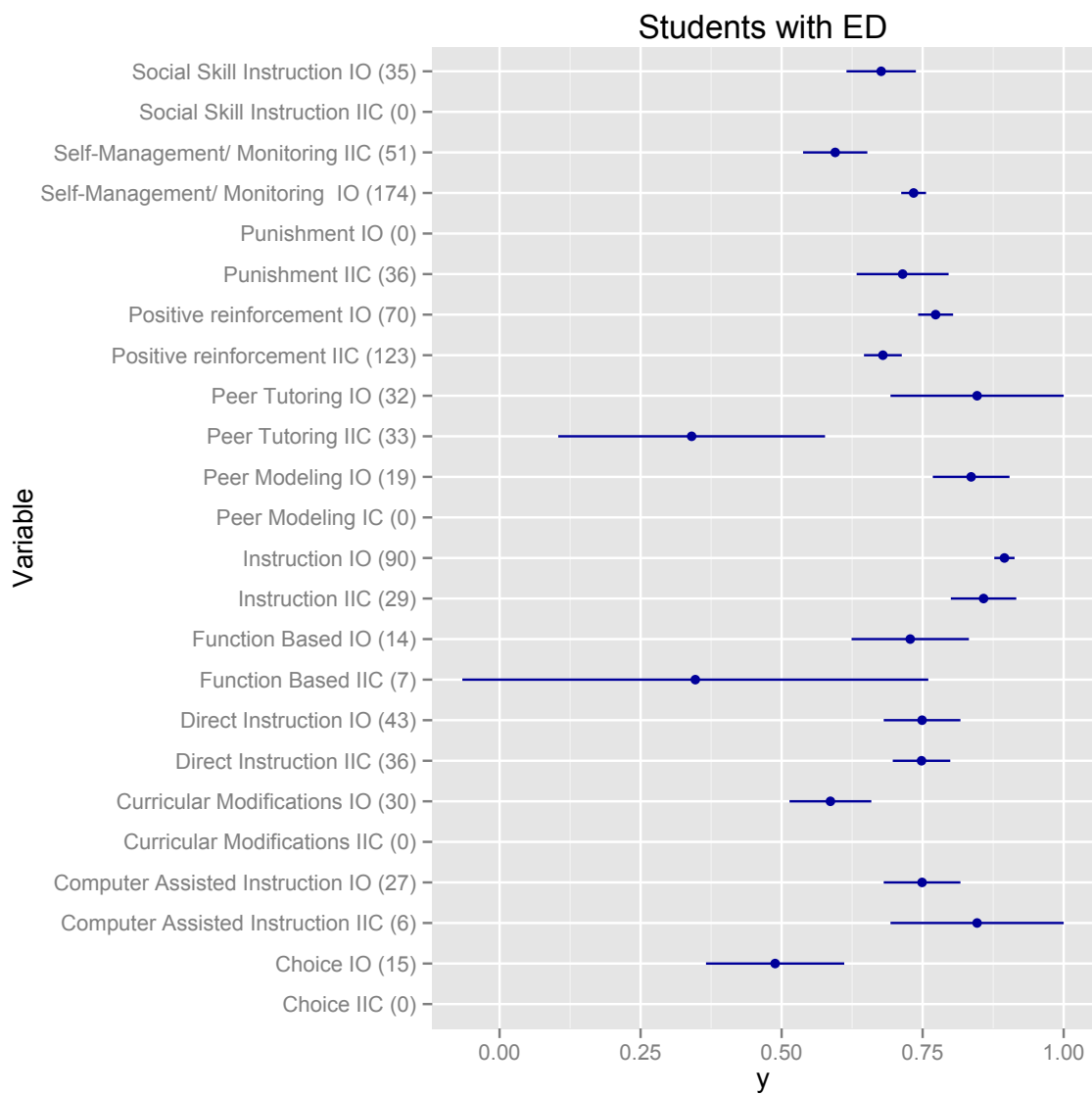
*Note:* RPhi with CI= 95% confidence interval.

Figure 7 Effect Size Comparison by Primary Study Quality



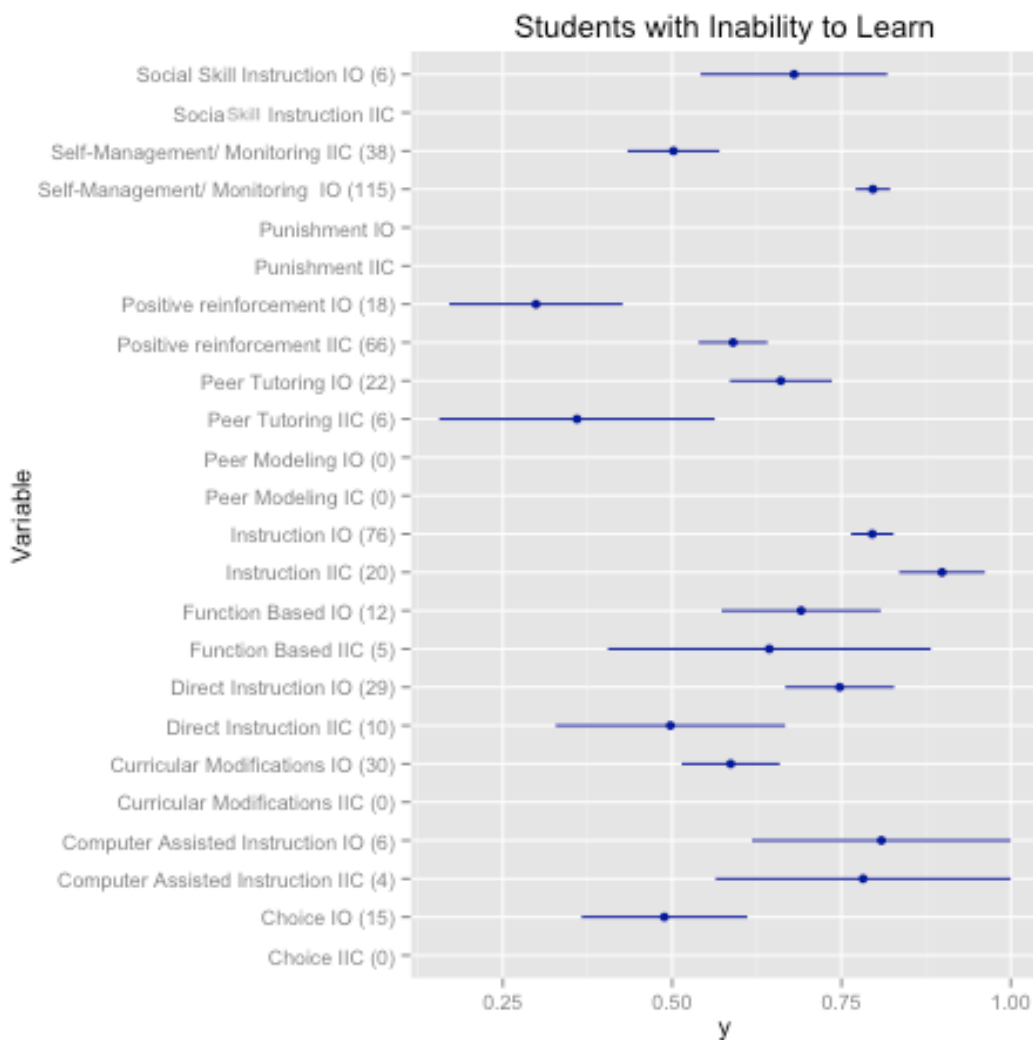
Note: RPhi with CI= 95% confidence interval.

Figure 8 Effect Size Comparisons of Interventions Used with Students with ED



*Note:* RPhi with CI= 95% confidence interval.

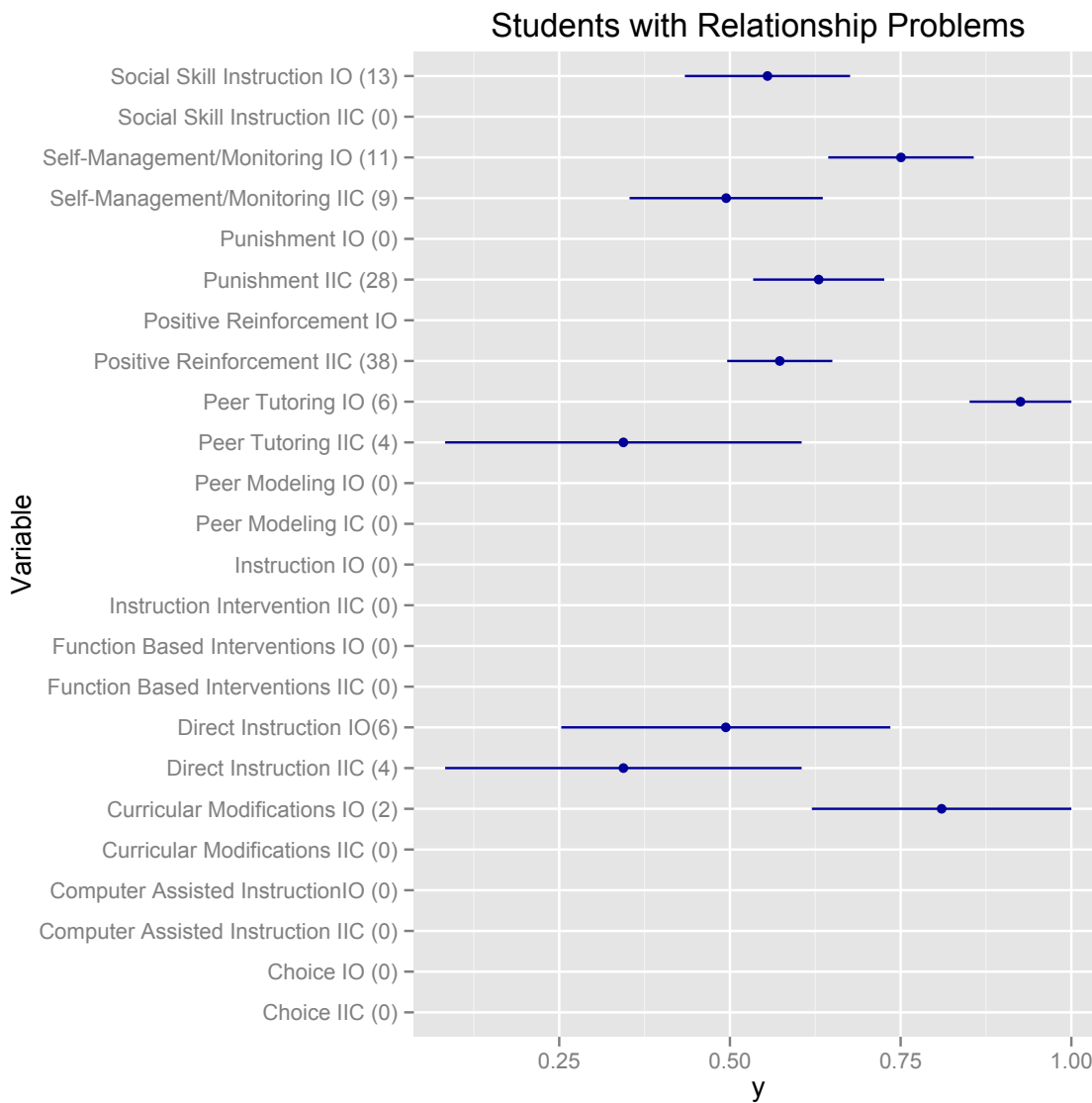
Figure 9 Effect Size Comparisons of Interventions Implemented with Students with the Inability to Learn Characteristic <sup>1</sup>



Note: RPhi with CI= 95% confidence interval.

<sup>1</sup> Participants may have also been reported as having other ED characteristics in addition to Inability to Learn.

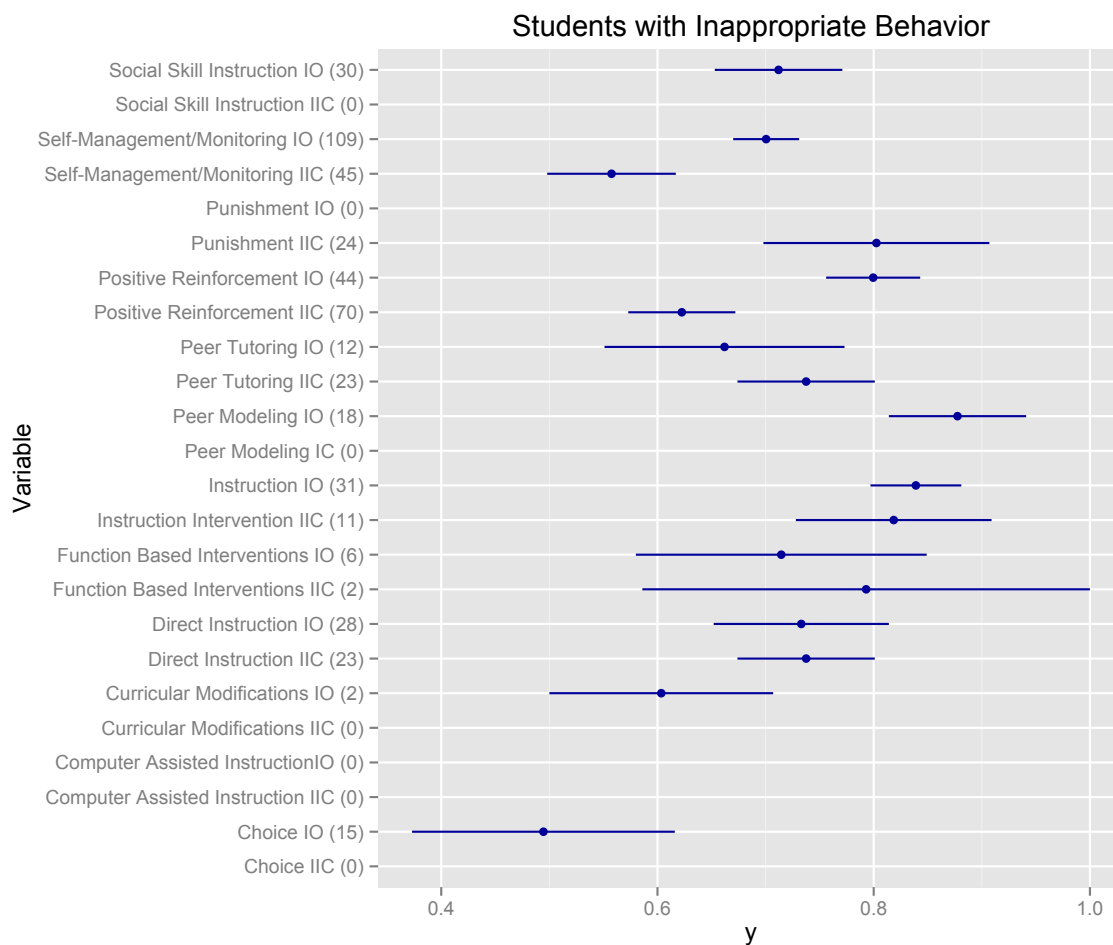
Figure 10 Effect Size Comparisons of Interventions Implemented with Students with Relationship Problems <sup>1</sup>



Note: RPhi with CI= 95% confidence interval.

<sup>1</sup> Participants may have also been reported as having other ED characteristics in addition to Relationship Problems.

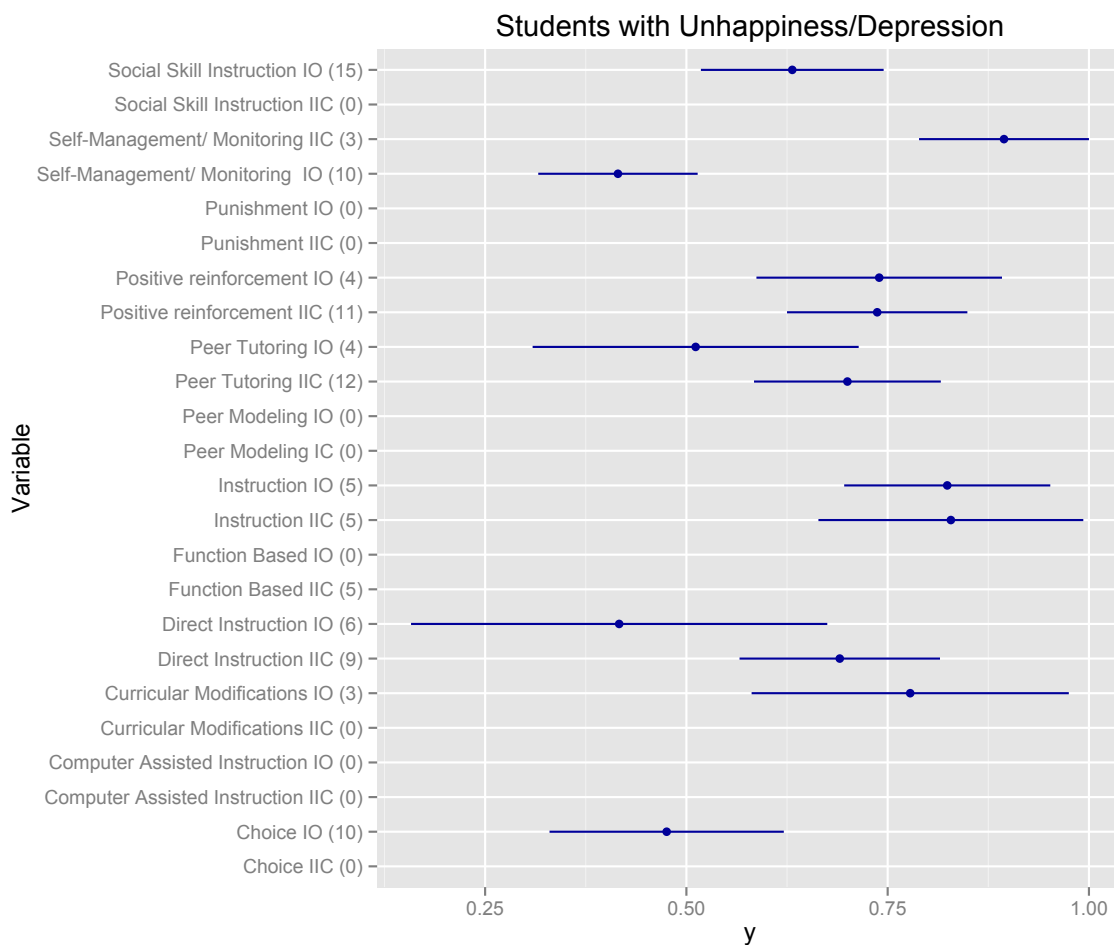
Figure 11 Effect Size Comparison of Interventions Implemented with Students with Inappropriate Behavior <sup>1</sup>



*Note:* RPhi with CI= 95% confidence interval.

<sup>1</sup> Participants may have also been reported as having other ED characteristics in addition to Inappropriate Behavior.

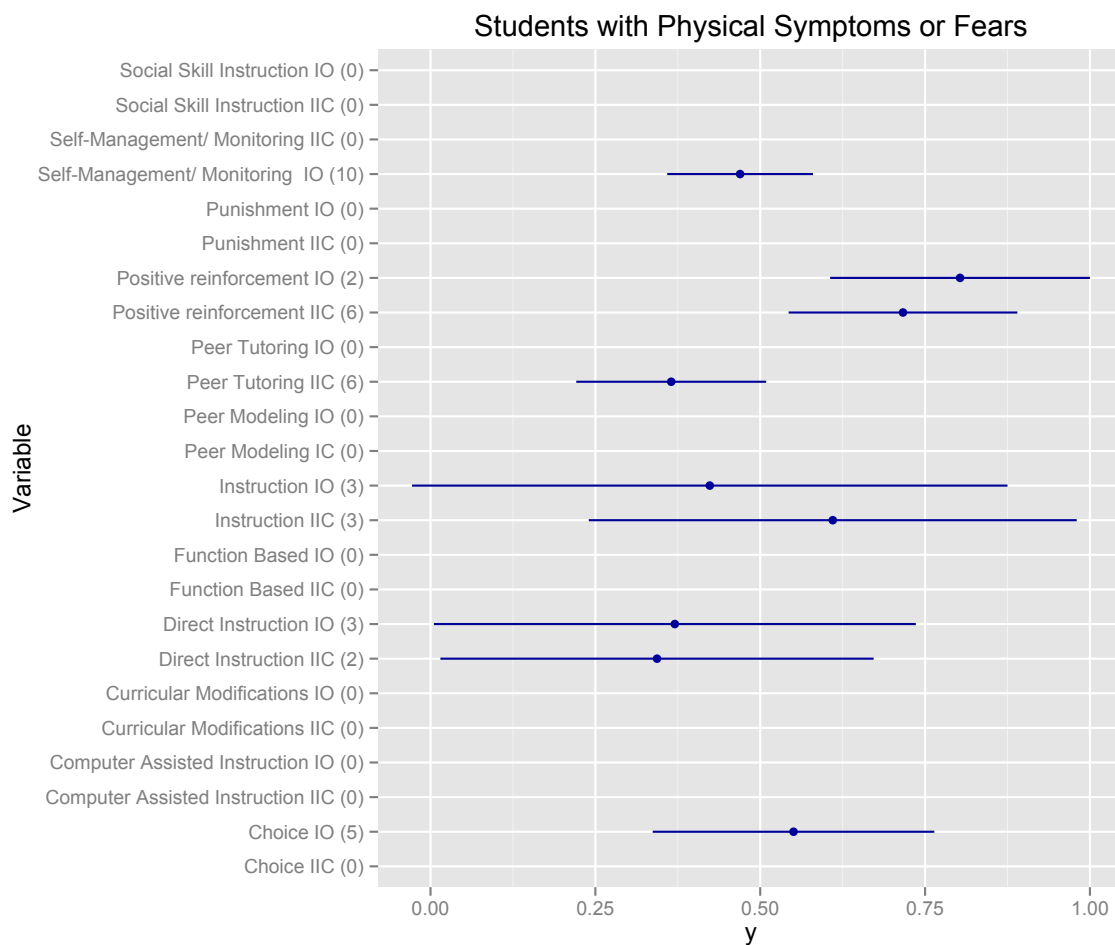
Figure 12 Effect Size Comparisons of Interventions Implemented with Students with Unhappiness or Depression <sup>1</sup>



Note: RPhi with CI= 95% confidence interval.

<sup>1</sup> Participants may have also been reported as having other ED characteristics in addition to Unhappiness or Depression.

Figure 13 Effect Size Comparisons of Interventions Implemented with Students with Physical Symptoms or Fears <sup>1</sup>

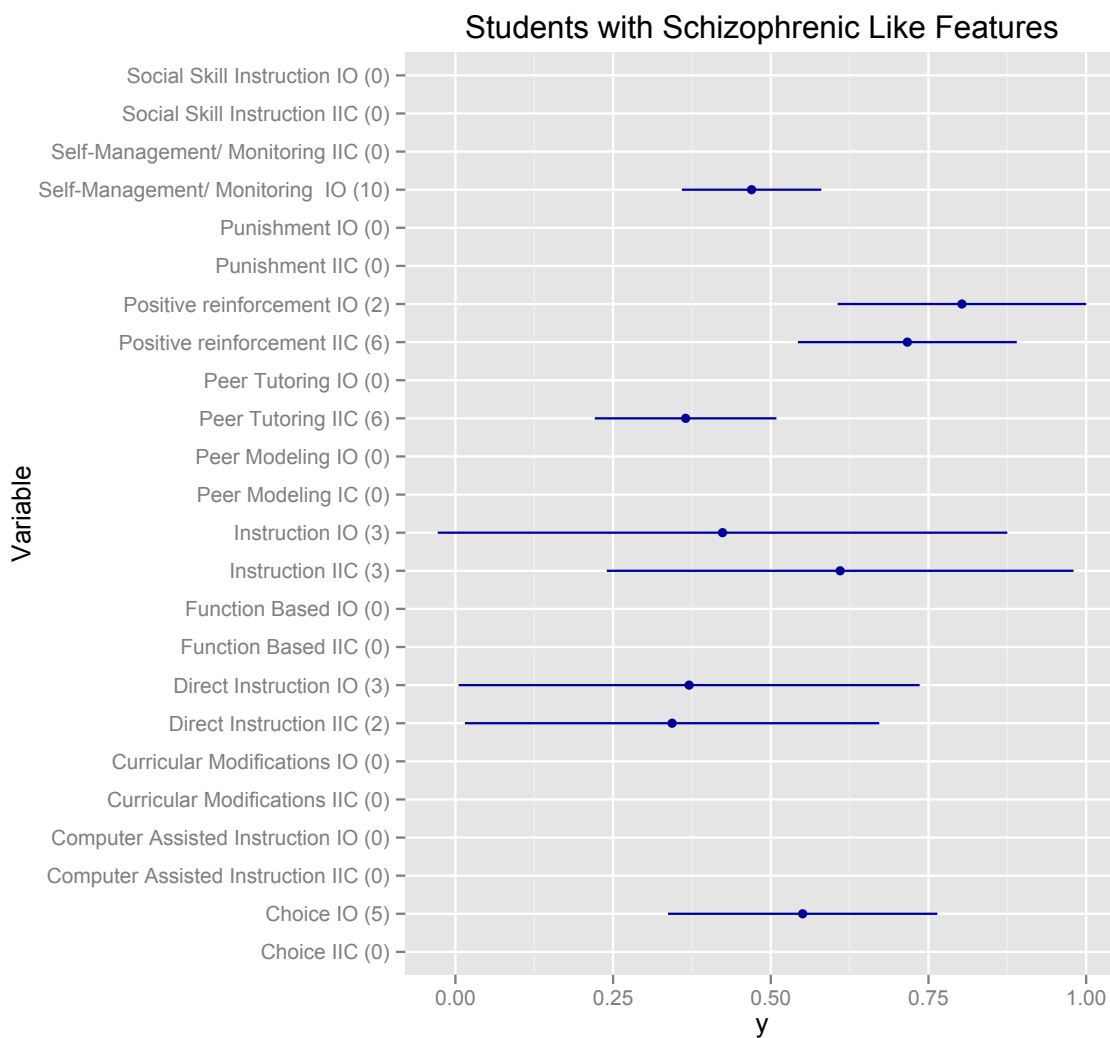


Note: RPhi with CI= 95% confidence interval.

<sup>1</sup> Participants may have also been reported as having other ED characteristics in addition to Physical Symptoms or Fears.



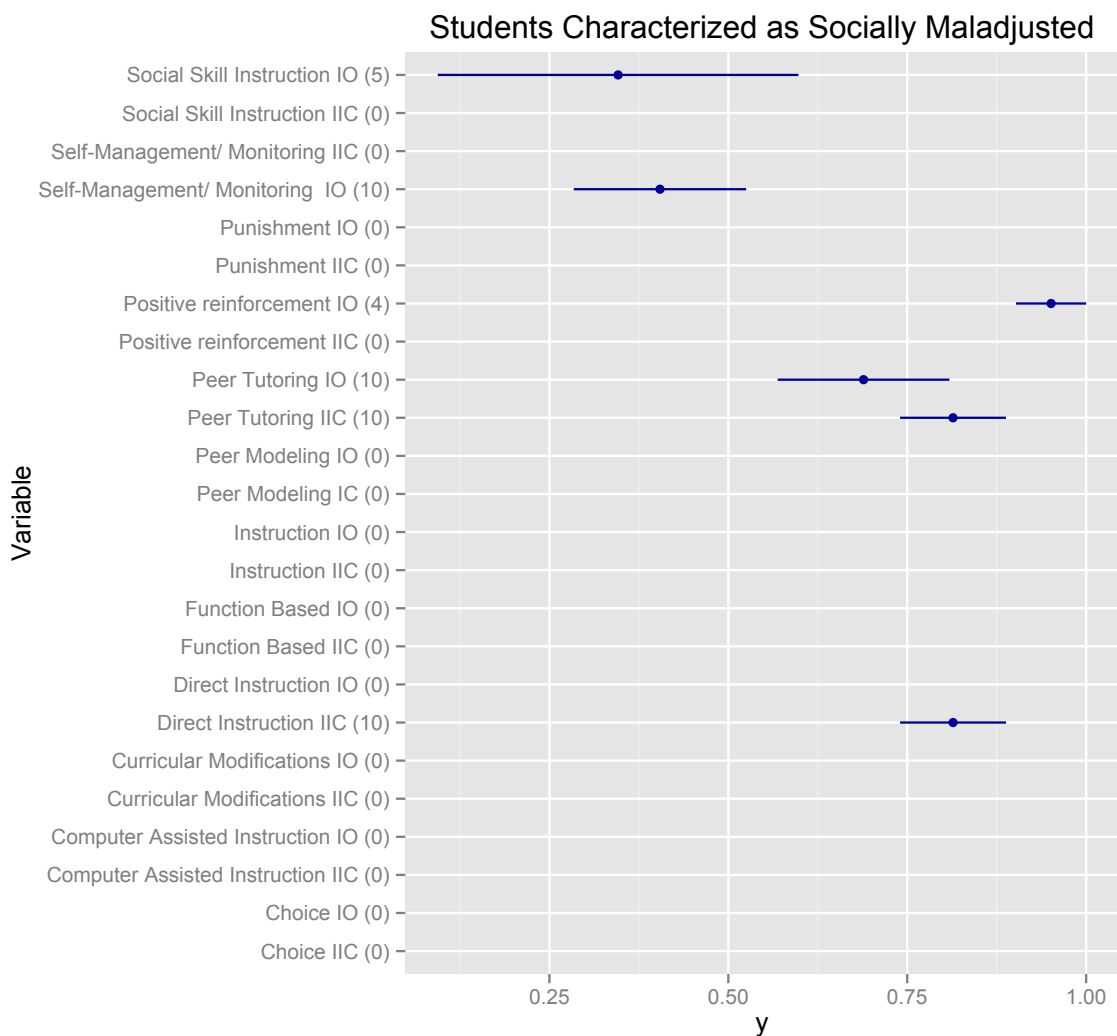
Figure 14 Effect Size Comparisons of Interventions Implemented with Students having Schizophrenia-Like Features <sup>1</sup>



Note: RPhi with CI= 95% confidence interval.

<sup>1</sup> Participants may have also been reported as having other ED characteristics in addition to Schizophrenia-Like Features.

Figure 15 Effect Size Comparisons of Interventions Implemented with Students characterized as Socially Maladjusted <sup>1</sup>



*Note:* RPhi with CI= 95% confidence interval.

<sup>1</sup> Participants may have also been reported as having other ED characteristics in addition to being Socially Maladjusted.