LANGUAGE FUNCTIONING AND YOUTH WITH CONDUCT PROBLEMS: A META-ANALYSIS

Cameron Kirkland Collins

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the School of Education (School Psychology).

Chapel Hill
2006

Approved by
Stephen R. Hooper, Ph.D.; Chair/Adv
John C. Brantley, Ph.D.
Barbara H. Wasik, Ph.D.
William B. Ware, Ph.D.
Mitchell J. Prinstein, Ph.D.

ABSTRACT
CAMERON K. COLLINS: Language Impairment in Youth with Conduct Problems: A Meta-Analysis
(Under the direction of Stephen Hooper, Ph.D.)

In an effort to better understand conduct problems among children and adolescents, considerable research has focused on the neuropsychological characteristics of youth with such problems. Language, one of several neuropsychological constructs, has been linked to conduct problems in youth. However, there remain many unanswered questions regarding this relationship. Therefore, this study seeks to quantitatively describe the association between conduct problems and language function using meta-analytic procedures. Analyses are guided by the following research questions: What is the magnitude of mean effect size for global language functioning in youth with conduct problems? Do studies evidence a relationship between conduct problems and more specific constructs of language function such as receptive and expressive language? How do certain variables (e.g., age, gender, ethnicity, presence of Attention-Deficit/Hyperactivity Disorder, and type of conduct problem) impact the relationship between conduct problems and language functioning? Relevant articles were identified by keyword searches of the Psych INFO database as well as by examining reference lists of collected articles. This process identified 235 contemporary research articles (i.e. conducted since 1980), which were reviewed for relevant information and subjected to inclusion criteria. Seventeen studies met the following criteria for inclusion: empirical studies conducted since 1980, employing group contrast design with non-disordered controls to investigate language functioning in participants younger than 21-years

of age, and utilizing standardized measures of language function. Three separate meta-analyses, one for each language construct, investigated the distribution of standardized mean difference effect size statistics (Hedges's *g*). Analysisalso include d heterogeneity testing and moderator analysis. Results indicated significant effect sizes for global, receptive, and expressive language in the moderate to strong range. Findings also suggested that ethnicity moderates the relationship between language functioning and conduct problems, with minorities at greater risk than non-minorities. Overall, findings provided strong evidence for an underlying neuropsychological deficit in language functioning in many youth with conduct problems, with some demographic variables moderating the magnitude of these effects.

ACKNOWLEDGEMENTS

I would like to thank Phil Icard, M.A. for assisting me in this work. Also, I would like to thank Stephen Hooper, Ph.D. for his guidance throughout this process. Finally, I thank my husband and parents for their support.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ADD……...Attention Deficit Disorder

ADHD……Attention-Deficit/Hyperactivity Disorder

AL…….....Adolescent Limited (delinquency)

APA……..American Psychological Association

ASB……Anti-Social Behavior

BED…….…….Behavioral-Emotional Disturbance

CD……...……Conduct Disorder

CELF……....Clinical Evaluation of Language Fundamentals

CELF-R……..Clinical Evaluation of Language Function-Revised

CMA……..Comprehensive Meta-Analysis

CP……....Conduct Problem

DABS…….…..Devereux Adolescent Behavior Scale

DTLA…….Detroit Test of Learning Aptitude

DSM……..……Diagnostic and Statistical Manual of Mental Disorders

DSM-IV……...Diagnostic and Statistical Manual of Mental Disorders-Fourth Edition

DSM-IV-TR……Diagnostic and Statistical Manual of Mental Disorders–Fourth Edition - Text Revision

EBD…….……Emotionally and Behaviorally Disturbed/Disabled

EOWPVT……...……..Expressive One-Word Picture Vocabulary Test

EOWPVT-UX……Expressive One-Word Picture Vocabulary Test- Upper Extension

FKSB……....Florida Kindergarten Screening Battery

FSIQ……..Full Scale Intelligence Quotient

IDEA……...Individuals with Disabilities Education Act

IQ……...Intelligence Quotient

LCP……...Life Course Persistent (delinquency)

LI……..Language Impairment

LNNB……Luria Nebraska Neuropsychological Battery

MFFT……Matching Familiar Figures Test

MMPI…….….Minnesota Multiphasic Personality Inventory

ODD……...….Oppositional Defiant Disorder

PIAT……..Peabody Individual Achievement Tests

PINTS……..Pittsburgh Initial Neuropsychological Test System

PIQ……Performance Intelligence Quotient

PPVT……..Peabody Picture Vocabulary Test

SCOLP……..Speed and Capacity of Language Processing

SD…….Standard Deviation

SED…….Seriously Emotionally Disturbed

SES…….....Social Economic Status

SMD……Standardized Mean Difference

TLC…….Test of Language Competence

TLC-E……....Test of Language Competence-Expanded

TOAL……...Test of Adolescent and Adult Language

TOLD……Test of Language Development

TOLD-I……Test of Language Development-Intermediate

TBI……..Traumatic Brain Injury

VADS……....Visual Aural Digit Span

VIQ…….……Verbal Intelligence Quotient

WAIS…….Wechsler Adult Intelligence Scale

WAIS-R……..Wechsler Adult Intelligence Scale-Revised

WCST……Wisconsin Card Sorting Test

WPPSI-R…….Wechsler Preschool and Primary Intelligence Scale-Revised

WRAT……Wide Range Achievement Test

WISC……..Wechsler Intelligence Scale for Children

WISC III……..…Wechsler Intelligence Scale for Children-Third Edition

WISC R…….Wechsler Intelligence Scale-Revised

CHAPTER I

INTRODUCTION

Recent increase in media attention surrounding youth with conduct problems has given rise to mounting concern about youth conduct problems. Some consider this a major public health problem (Farrington & Loeber, 2000; Essau, 2003), as youth with conduct problems negatively impact society. In addition to the direct negative influence on other individuals, youth with conduct problems affect society via more indirect routes. Lambert and colleagues (2001) evaluated the monetary costs incurred by youth with Conduct Disorder (CD), looking specifically at the cost of mental health treatment for children with CD and those without CD. The cost of treatment for a child with CD was significantly greater than the cost of treatment for a child without CD; with mean costs of $21,000 and $8,000, respectively (Essau, 2003). Considering that children and adolescents with conduct problems account for up to one-half of all psychiatric referrals (Kazdin, 2000; Webster-Stratton, 1993), it is reasonable to infer that the mental health system is substantially overburdened. The chronic involvement of youth with conduct problems in the judicial system is a second source of financial burden. Cohen (1998) reported that the cost of law enforcement, process of adjudication, and incarceration incurred by one juvenile delinquent engaged in four years of juvenile offending and ten years of adult offending ranged from 1.7 million to 2.3 million dollars (Farrington & Loeber, 2000).

Conduct Problems: Nomenclature and Taxonomies

Several labels fall under the broader umbrella of "conduct problems." Although children and adolescents labeled Oppositional Defiant Disordered (ODD), Seriously Emotionally Disturbed (SED), behaviorally disordered, and delinquent share many of the same characteristics, these labels are associated with different contexts. CD, ODD, and SED are formal labels of classification whereas "delinquent" and "behaviorally disordered" are used more informally to describe a person or their behavior. The labels CD and ODD are psychiatric diagnoses described in the Diagnostic and Statistical Manual of Mental Disorders Fourth Edition-Text Revision (DSM-IV-TR) and are primarily used in psychiatric and mental health contexts. SED is a federally defined special education label used within school systems. The use of the descriptive label "behaviorally disordered" is not restricted to any specific context and is used as a general descriptor for a pattern of externalizing behavior problems. The term "delinquent" is used within the judicial system to describe youth who commit crimes. The term has also made its way into the everyday vernacular of the public. These different terms, their overlap, and the subtle differences between them present significant challenges to researchers as they do not always mean the same, nor are they mutually exclusive.

One way practitioners typically address this problem is by lumping all these terms together under the broader umbrella of "conduct problems." Although this may simplify things conceptually, using the term "conduct problems" wrongly implies homogeneity and such generalization hinders the progress of research. As this is becoming increasingly clear, the primary goal of much research in this area is to delineate and understand the characteristics of youth with conduct problems, and to develop a nomenclature that accurately depicts the differences between children who display different constellations of

problem behavior. Only in this fashion can the field move toward evidence-based diagnostic processes and treatments.

Conduct Disorder (CD)

McMahon and Wells (1998) describe CD as "a recurrent, persistent pattern of behavior in which the child violates the basic rights of others or major age-appropriate societal norms or rules" (p.112). The DSM-IV-TR (American Psychological Association, 2000) also provides formal diagnostic criteria for CD (see Appendix A). Compared to some of the other diagnoses in the DSM-IV-TR, "CD is rather different because it consists of a group of behaviors, none of which is conceptually central to our understanding of the disorder. The only requirement is that individuals should manifest a lot of these behaviors if they are to be given the diagnosis. Even at the level of conceptual grouping, the items constituting CD are not immediately and self-evidently coherent" (Angold & Costello, 2001, p. 126).

The DSM-IV-TR further subdivides CD into Childhood Onset or Adolescent Onset. The Childhood Onset specifier is reserved for youth demonstrating sufficient characteristic behaviors before the age of 10 years and indicates a greater degree of pathology. Children with this subtype exhibit more serious problem behavior, which often persists and is not easily rehabilitated by intervention. For these reasons, Childhood Onset CD is associated with poor prognosis (Farrington & Loeber, 2000; Kaufman, 2001). Adolescent Onset CD develops after a relatively normal childhood, consists of less serious offenses, and has a better prognosis (Farrington & Loeber, 2000). Estimates of prevalence rates for CD range from less than 1% to 16% (Essau, 2003: Loeber, Burke, Lahey, Winters, & Zera, 2000). When considering the prevalence rate for CD in clinical settings, the rate increases to 26%

(Essau, 2003). CD is almost three times more likely to occur in boys than girls (APA, 2000; Loeber, et. al., 2000).

Oppositional Defiant Disorder (ODD)

ODD, also a psychiatric diagnostic label, describes a recurrent pattern of behavior in which a child demonstrates defiance, disobedience, negativity, and hostility toward authority figures (Alvarez & Ollendick, 2003; APA, 2000; McMahon & Wells, 1998). Appendix A shows the DSM-IV-TR diagnostic criteria (APA, 2000) for ODD. In general, some degree of oppositionality and defiance is expected during early childhood and, to a certain extent, during adolescence. When considered within specific contexts, such as a young child attempting to establish independence from their parents, opposition and defiance are developmentally appropriate. These behaviors constitute a clinical problem when they occur at a much greater frequency compared to other children at the same developmental level and age, and when behavior interferes with or impairs the child's ability to function in some aspect of his or her daily life. Loeber et al. (2000) reviewed prevalence rates of ODD in seven studies conducted between 1987 and 1998. They found prevalence rates ranging from 1.5% to 15.6%; and, as with CD, ODD occurs more frequently among boys than girls.

Serious Emotional Disturbance (SED)

A third classification of conduct problems, SED, is particularly pertinent to school psychologists and other professionals in education. Depending on state-to-state differences, SED is also referred to as Behavioral-Emotional Disturbance (BED) and Emotionally-Behaviorally Disturbed/Disabled (EBD). The federal definition according to Individuals with Disabilities Education Act-Part B (IDEA-Part B) defines SED as follows:

> 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 educational performance: an inability to build or maintain satisfactory interpersonal relationships with peers and teachers; inappropriate types of behavior or feelings under normal circumstances; a general pervasive mood of unhappiness or depression; or a tendency to develop physical symptoms or fears associated with personal or school problems. The term also includes schizophrenia. The term does not apply to children who are socially maladjusted, unless it is determined that they have a serious emotional disturbance (Jacob-Timm & Hartshorne, 1998, p. 112).

As can be seen from the federal definition, children classified as SED represent a very heterogeneous group. Children in this group may have mood, anxiety, or psychotic disorders in addition to problems with their behavior. The prevalence of SED is difficult to establish because of state-to-state differences in eligibility requirements. The United States Office of Special Education Programs (1997) estimated that up to 10 million children suffer conditions that would qualify them for services under the classification of SED, but only 440,000 of these youth receive services through special education (Johnson-Reid, Williams, & Webster, 2001).

Delinquency

The term "delinquent" refers to a young person (usually an adolescent) who has committed a serious crime for which he or she is penalized by the criminal justice system. A young person may be considered delinquent for a wide range of behaviors: assault, theft, homicide, substance abuse, rape, destruction of property, possession of illegal substance, truancy, repeated traffic violations, etc. The 1999 National Report found a 35% increase in juvenile arrests between 1988 and 1997 (Essau, 2003). Looking specifically at documented violent juvenile crime (i.e., homicide, rape, aggravated assault, and robbery), prevalence rates increased 92% between 1987 and 1997 (Farrington & Loeber, 2000). These percentages represent arrests or convictions, not actual crimes committed; therefore, it is likely these statistics underestimate the full extent of the problem.

Moffitt and colleagues have conducted a multitude of studies investigating delinquency in a longitudinal birth cohort in New Zealand (Moffitt, 1990a; Moffitt, 1993b; Moffitt, Lynam, & Silva, 1994). This line of research seeks to "delineate possible subtypes among delinquents and to examine their correlates, developmental precursors, and ultimate outcomes" (Moffitt, 1990b, p.893). Their research findings suggest two subtypes of delinquency: Life Course Persistent (LCP) and Adolescent Limited (AL). LCP offenders tend to be recognized early in childhood, are responsible for a disproportionately large percentage of delinquency, and continue to commit crimes into adulthood (Moffitt, 1993b). Individuals who are described as LCP delinquents frequently demonstrate clinically significant neuropsychological deficits (Moffitt et al., 1994). AL delinquency begins in adolescence and usually does not persist past this age period (Moffitt, 1993b). This subtype is less often associated with neuropsychological deficits (Moffitt, 1993b). These subtypes closely parallel the DSM-IV-TR psychiatric diagnoses of Child Onset and Adolescent Onset CD.

Continuity of Conduct Problems

Many studies investigating the development and course of conduct problems emphasize the continuity and stability of disruptive behavior (Caspi, Elder, & Bem, 1987; Farrington & Loeber, 2000; Loeber, 1991; Loeber, Green, Keenan, & Lahey, 1995; Tremblay, Phil, Vitaro, & Dobkin, 1994). Nearly 30 years ago, Olweus (1979) noted the strong correlation (.63) between early aggression and later aggression, which approximates the stability of intelligence over time (Loeber & Coie, 2001). According to the diagnostic histories of youth with CD, as many as 80-90% previously carried a diagnosis of ODD (Loeber et al., 1995; Loeber, 1988). Other studies provide additional support for the

predictive utility of ODD in anticipating later CD (Alvarez & Ollendick, 2003; Lahey & Loeber, 1994).

Continuity in problem behavior is also found among delinquent youth. Several studies demonstrate that early antisocial behavior is one of the strongest predictors of later delinquency (Loeber & Dishion, 1983; Loeber & Stouthamer-Loeber, 1987). Because of the stability and persistence of problem behavior over time, the importance of early identification, problem delineation and correlates, and intervention cannot be overstated.

Theoretical Conceptualization: The Linkage of Language with Conduct Problems

Many researchers approach the study of conduct problems from a neuropsychological perspective. Studies assessing the neuropsychological characteristics of children with conduct problems seek to identify a link between various aspects of neurocognitive functioning and problem behavior, and to describe the nature of such a relationship. Specific neuropsychological constructs of interest include: general cognitive ability (i.e. intelligence), memory, executive functioning, sensory perception, motor functioning, and language. While all of these constructs have been linked to conduct problems in one study or another, the current investigation focuses on one neuropsychological construct, language, which has frequently been related to conduct problems in children and adolescents.

Neuropsychological linkages to conduct problems. In examining the relationship between neuropsychological deficits (e.g. language impairment) and conduct problems, three hypotheses regarding causality must be considered. The first hypothesis posits that conduct problems lead to neuropsychological deficits; however, no studies provide evidence of conduct problems preceding neuropsychological deficits. The second hypothesis proposes that the direction of the relationship runs from neurological deficit to later conduct problems.

This hypothesis has received the most consistent support (Ayllon & Roberts, 1974; Moffitt, 1990b, 1993a; Schonfeld, Shaffer, O'Connor, & Portnoy, 1988). The third hypothesis suggests that a third factor contributes to the development of both conduct problems and neuropsychological deficits. Some research supports this hypothesis of common antecedents (Huesmann & Eron, 1984; Olweus, 1979) while other research does not (Lynam, Moffitt, & Stouthamer-Loeber, 1993; Schonfeld et al., 1988; Sobotowicz, Evans, & Laughlin, 1987).

Lynam and Henry (2001) examine the issue of mechanism in a review of existing research. They concluded that although the "evidence suggests the causal direction runs from poor neuropsychological functioning to serious conduct problems;this conclusion cannot be drawn unequivocally" (p. 256). This is largely because not all children with neuropsychological deficits (e.g., language deficits) will show conduct problems. In order to unequivocally demonstrate this direction of causality, "three conditions must be met: (1) neuropsychological problems must be positively related to [Anti-Social Behavior (ASB)], (2) neuropsychological problems must precede the ASB, and (3) it must be possible to rule out plausible alternative explanations of the relation" (Lynam & Henry, 2001, p. 236). Although many studies have addressed these conditions, no single study has successfully met all three. Regardless of the direction, the examination of neuropsychological deficits as being related to conduct problems remains a strongly viable area of investigation.

Language linkages to conduct problems. Many researchers have identified language impairment as a crucial neuropsychological deficit potentially leading to later problems in various areas of functioning. For children at-risk for conduct problems, the importance of competence in communication has been established as language is the primary means of establishing and maintaining successful relationships, constitutes a means of organizing

behavior, and is central to the successful acquisition of many cognitive and academic skills (Warr-Leeper, Wright, & Mack, 1994). Some theories provide a proximal explanation and others approach the issue from a more distal, developmental perspective.

Proximal theories propose a direct link between language deficits and conduct problems. For example, problem behavior may serve a communicative function for many children with verbal difficulties (Davis, Sanger, & Morris-Friehe, 1991; Denno, 1986; Humber & Snow, 2001; Sager, Hux, & Ritzman, 1999; Sanger, Creswell, Dworak, & Schultz, 2000; Warr-Leeper, Wright, & Mack, 1994; Wickstrom-Kane & Goldstein, 1999). Children who have difficulty with verbal communication often use alternative, nonverbal means of communicating. For example, a child with language impairment may grab a toy from another child's hands because the child with language impairment is not able to verbally ask to play with the toy. Or, a child with language impairment may behave aggressively as a way of expressing frustration or anger. As previously noted, this type of behavior is developmentally appropriate at young ages when children have not yet achieved communicative mastery. This typical behavior becomes a problem when the young child grows older and still lacks the language abilities necessary for appropriate interactions. Thus, these alternative means of communicating may continue to be utilized in social interactions. For young children with language difficulties, problem behaviors may be the only available option for effective communication. Unfortunately, aggressive, disruptive, and coercive behaviors tend to result in their desired outcome and, consequently, are reinforced and maintained (Humber & Snow, 2001; Speltz, DeKlyen, Calderon, Greenberg, & Fisher, 1999; Wickstrom-Kane & Goldstein, 1999). In this fashion, proximal theories fail to account for the increase in severity of ASB across time (Loeber, 1988).

Other theories take a more distal and developmental perspective in explaining the connection between language ability and other areas of functioning such as social functioning, behavioral and emotional self-regulation, and academic success. According to such theories, language impairment is indirectly linked to later conduct problems via moderating variables. Moderator variables are those that affect the direction and/or strength of the relationship between the predictor and outcome variables; moderators establish "when" or "for whom" a variable most strongly predicts an outcome (Frazier, Tix, & Barron, 2004). This is in contrast to mediating variables, which explain the mechanism through which a predictor influences an outcome variable; mediators establish "how" or "why" a variable predicts an outcome (Frazier, Tix, & Barron, 2004). Lynam and Henry (2001) emphasize that normal language ability is a crucial ingredient for prosocial behavior including delayed gratification, anticipation of consequences, and linking belated punishment with previous misbehavior. Normal language ability is also necessary for successful social-information processing (Wong & Cornell, 1999). Deficits in language potentially lead to deficits in other areas of social functioning which, in turn, contribute to deviant behavior.

Vygotsky's social development theory (1962) provides useful insight into the importance of language in serving a self-regulatory function. This process begins in early childhood when children use self-talk (verbalized thought) to regulate their behavior. As their language competence increases over time, their self-regulatory skills strengthen and the child no longer needs to verbalize their thinking; and instead, the verbalized thought becomes internalized. This internalized thought takes over the role of verbalized thought in behavioral self-regulation. Luria (1966) noted that this verbal control over behavior begins to emerge around age 3.5 years.

Contemporary research addresses the role of language in behavioral self-regulation. Verbal deficits may impair a child's capacity to develop adequate internalized verbal thought, which is necessary for inhibitory, behavioral self-regulation (Alvarez & Ollendick, 2003; Buikhuisen, Bontekoe, Plas-Korenhoff, & Meijs, 1988; Tarter, Hegedus, Winsten, & Alterman, 1984; Yeudall, 1980). Behavioral regulation allows for delayed gratification, anticipation of consequences, and planning (Alvarez & Ollendick, 2003). Wilson and Herrnstein (1985) suggest that language plays an essential role in linking behavior with its consequences, especially delayed consequences. Consequently, problems with internalized speech can hinder a child's problem solving capacity (Wilson & Herrnstein, 1985).

Contemporary research also addresses the role of language in emotional self-regulation and suggests that verbal deficits contribute to difficulties with self-regulation of emotion (Alvarez & Ollendick; 2003; Cole, Usher, & Cargo, 1993; Cook, Greenberg, & Kusche, 1994; Speltz et al., 1999). Children with poor affective-state vocabulary (i.e. words to express feelings) have limited ability to understand, verbally express, and regulate their emotions (Speltz et al., 1999). Matching emotion words with nonverbal emotional expression requires adequate verbal ability (Cole, Usher, & Cargo, 1993). Cook, Greenberg and Kusche (1994) found that children with behavior problems experienced difficulties in verbalizing their emotional experiences and identifying emotional cues in themselves and others. Because of these limitations in verbal expression, such children may be more likely to engage in problem behaviors.

Some hypothesize that language problems lead to conduct problems by way of school failure (Buikhuisen et al., 1988; Hirschi & Hindelang, 1977; Meltzer, Roditi, & Fenton, 1986; Moffitt, Gabrielli, Mednick, & Schulsinger, 1981). Most children and adolescents

receive education in a public school setting, an atmosphere with substantial verbal demands. A child must possess and utilize sufficient verbal skills to succeed in academics and to appropriately negotiate the social context (Buikuisen et al., 1988; Humber & Snow, 2001). Children with poor verbal skills frequently fail to achieve the rewards of academic success and peer belongingness. Initial experiences of failure and frustration in school may contribute to later delinquency in many ways: damaging self-esteem, restricting possible future opportunities, creating a negative attitude toward authority, peer rejection and alienation, and subsequently leading to association with a deviant peer groups (Buikhuisen et al., 1988; Meltzer, Roditi, & Fenton, 1986; Moffitt, Gabrielli, Mednick, & Schulsinger, 1981).

Assessing Language Problems in Youth with Conduct Problems

Researchers and clinicians assess language using a variety of methods. Numerous studies use Verbal IQ (VIQ) as an indicator of global languageability. Wechsler's intelligence scales are commonly used in research regarding children with conduct problems. This family of standardized intelligence measures possesses a psychometric and conceptual continuity that has provided practitioners and researchers with a common understanding regarding measurement of cognitive abilities that applies to people of all ages. Studies investigating WISC-IV scores in children with expressive language disorders and mixed expressive-receptive language disorders showed that effect sizes based on group mean composite scores were the largest for the verbal index (The Psychological Corporation, 2003).

Neuropsychological assessments typically cover a wider range of functions than intelligence measures. Neuropsychological assessment seeks to detect functional deficits by

using “performance measures designed to evaluate individuals along a continuous dimension of proficiency” (Gorenstein, 1990, p.30). Some researchers use well established assessment batteries, such as the Luria Nebraska Neuropsychological Test Battery or the Halstead Reitan Neuropsychological Test Battery. These batteries combine various tasks tapping a range of neuropsychological domains into one large standardized measure. Others create their own battery using a flexible or eclectic battery approach. Here, construct specific measures and subtests from larger instruments or batteries are selected to evaluate the particular constructs of interest. In general, most contemporary neuropsychological assessments include tasks/subtests that tap language-related abilities. Thus, studies investigating the neuropsychological status of youth with conduct problems are a potential source of rich information regarding the language abilities of this population.

The third approach used in evaluating the verbal abilities of children and adolescents with conduct problems involves assessment measures specifically designed to evaluate language functions. Such measures include the Clinical Evaluation of Language Fundamentals (CELF), Expressive One-Word Picture Vocabulary Test (EOWPVT), Peabody Picture Vocabulary Test (PPVT), Test of Adolescent and Adult Language (TOAL), Test of Language Competence (TLC), Test of Language Development (TOLD), and many others. These instruments are generally most familiar to Speech-Language Pathologists and are frequently used to diagnose language impairment. Compared to VIQ and some neuropsychological measures, these instruments provide the most thorough and specific analysis of an individual’s language abilities.

Tests designed for assessing language ability have the advantage of breaking down language functioning into specific components. At a general level, such measures provide

information on a child's receptive and expressive language abilities. Briefly stated, expressive language consists of the verbal expression of language (i.e., the ability to use verbal language to communicate needs, wants, intentions, and emotions); whereas receptive language consists of the comprehension of language (i.e., the ability to understand and process the spoken language of others). This conceptualization of language is consistent with descriptions found in the DSM-IV-TR (2000) regarding diagnoses of Expressive Language Disorder and Mixed Receptive-Expressive Language Disorder. Diagnostic criteria for these disorders are presented in Appendix A. Finally, language specific measures typically provide an index of "global" or "total" language functioning, which is a composite index comprised of the examinee's performance across all subtests. Of note, language-specific tests further divide these general domains into very specific language abilities such as phonological processing, receptive vocabulary, confrontational speech, spontaneous speech, fluency, pragmatics, and others; however, such specific skills are not the focus of the present investigation.

Certainly, these are not the only methods available for evaluating an individual's language abilities. Other methods include subjective rating scales, evaluation of language samples, and observation of social interaction. However, the current investigation focuses on standardized, quantitative measures with known validity and reliability.

Purpose and Hypotheses

The primary purpose of this study is to investigate the nature of language functioning in youth with conduct problems using meta-analytic procedures to synthesize the findings of previous research investigating language differences between youth with and without conduct problems. Although a critical review of this research suggests a relationship

between language difficulties and conduct problems, a meta-analytic approach can address this question with quantitative evidence. This study is important for several reasons. First, despite the abundance of research investigating language function and conduct problems, meta-analytic procedures have yet to be applied. Second, these procedures can detect different degrees of language difficulty in different "types" of conduct problems (e.g., CD versus ODD versus Delinquent). Such information can help to further delineate the nature of conduct problems. Third, this study can describe the nature of the relationship, if any, between specific language constructs (e.g., global language, receptive language, and expressive language) in youth with conduct problems. Fourth, meta-analytic procedures can identify moderating variables, potentially predicting which children are at the greatest risk for language difficulties. Fifth, clarifying the relationship between language dysfunction and conduct problems is important in terms of developing appropriate interventions.

Based on the available literature addressing the relationship between conduct problems and language functioning, variously defined, the following research questions and associated hypotheses will be examined:

Question 1. What is the magnitude of mean effect size for global language functioning in youth with conduct problems compared to youth without conduct problems?

Hypothesis 1. Given the available literature, it is suspected that meta-analytic findings will show a significant moderate effect size for global language functioning in youth with conduct problems.

Question 2. Will studies examining the linkages between conduct problems and language functions evidence any relationship with regard to the specific constructs of receptive and expressive language?

Hypothesis 2. When looking specifically at receptive and expressive language, it is suspected that effect sizes for these two language domains will be significant, and of comparable magnitude.

Question 3. It is unclear how certain key variables (e.g., chronological age, gender, ethnicity, comorbid ADHD, and type of conduct problem) may influence effect size of global, receptive, or expressive language functioning in individuals with conduct problems. Therefore, analyses seek to clarify whether or not these variables moderate the magnitude of effect size for global, receptive, or expressive language functioning. Again, as Frazier et al. (2004) noted, moderator variables are those that affect the direction and/or strength of the relationship between the predictor and outcome variables; moderators establish "when" or "for whom" a variable most strongly predicts an outcome; not to be confused with mediating variables, which explain the mechanism through which a predictor influences an outcome variable and establish "how" or "why" a variable predicts an outcome.

Hypothesis 3. It is hypothesized that effect size magnitude for global language, receptive language, and expressive language will differ with regard to age, gender, ethnicity, comorbid Attention-Deficit/Hyperactivity Disorder, and type of conduct problem. Specific hypotheses for these variables are as follows:

Age. Given the developmental nature of language, it is suspected that younger participant samples will exhibit larger effect sizes than older participant samples for each of the three language constructs.

Gender. It is hypothesized that studies with higher percentages of males will show larger effect sizes than studies with lower percentages of males for each of the three language constructs.

Ethnicity. It is hypothesized that studies with lower percentages of Caucasian participants will show larger effect sizes than studies with higher percentages of Caucasian participants for each of the three language constructs.

ADHD. Several previous studies have found that youth with conduct problems and comorbid ADHD demonstrate greater language difficulties than youth with conduct problems who do not have ADHD. Given this, it is hypothesized that effect size increases in magnitude as percentages of participants with ADHD in study samples increase.

Type of conduct problem. Although significant mean effect sizes are anticipated for each of the different types of conduct problems (e.g., Delinquent, CD, ODD, SED), it is suspected that meta-analytic findings will show significant differences in effect size for language impairment across the different types of conduct problems. In this regard, it is expected that larger effect sizes will be found for CD and ODD, as these labels likely represent more homogenous populations given their standard operational definitions, compared to the labels Delinquent and SED/BD.

These research questions and hypotheses were derived from the vast body of existing literature in this area. While many of these questions have been addressed in previous investigations, findings often produce conflicting or inconclusive evidence such that the answers to these questions remain unclear. Although the vastness of this body of literature can be considered an obstacle in attempting to arrive at a coherent understanding, especially when studies seem to reach dissimilar conclusions, it is this very proliferation of research that permits the application of meta-analytic statistical procedures. In a field of study with a long history, characterized by changes in classification system for describing conduct problems in youth as well as increasing sophistication and specificity with regard to the measurement of

language functioning, meta-analytic procedures are necessary to sort through existing research findings. The following comprehensive review of this prolific body of literature will illustrate the evolving nature of this field of study, surely revealing many yet unanswered questions.

CHAPTER II

LITERATURE REVIEW

Review of Pre-1980 Research

Although the current investigation excludes research conducted prior to 1980 from statistical analysis, understanding the origins of this line of research helps to place the current study within a broader context. Because the child psychiatric diagnoses CD and ODD first emerged in the 1980 edition of the DSM, early research focused on delinquent populations.

Much of the early research in this line of investigation was ignited by the frequently cited assertion that delinquents were characterized by a VIQ < PIQ discrepancy on intelligence scales (Berman & Seigal, 1976; Culberton, Feral, & Gabby, 1989; Fernald & Wisser, 1967; Grace & Sweeney, 1986; Haynes & Bensch, 1981; Henning & Levy, 1967; Prentice & Kelly, 1963; Walsh & Beyer, 1986). This conclusion received consistent support from many studies prior to 1980 (Camp, Zimet, van Doorninck, & Dahlem, 1977; Hays, Solway, & Schreiner, 1978; Henning & Levy, 1967; Lewis, Shanok, Pincus, & Glaser, 1979; Manne, Kandel, & Rosenthal, 1962; Maskin, 1974; Ollendick, 1979; Solway, Hays, Roberts, & Cody, 1975; West & Farrington, 1973). In an earlier review of this literature, Prentice and Kelly (1963) examined 24 studies that investigated the IQ scores (as measured by Wechsler scales) in delinquent populations. They noted:

> Almost without exception, these studies based largely on an adolescent population report the significant elevation of Performance over Verbal IQs. Moreover, this pattern is sustained generally in the majority of other studies in spite of the decided variations in age, sex, race, setting, and form of Wechsler scale administered, as well as substantial differences between the criteria for delinquency (p. 333).

Some evidence did not support this conclusion, or at the least introduced doubt to this claim. Fernald and Wisser (1967) examined WISC IQ scores in a group of adolescent male delinquents to determine whether the magnitude of VIQ < PIQ discrepancy predicted degree of delinquent behavior as indicated by police records. Results of the analysis between VIQ < PIQ discrepancy and degree of acting out indicated a non-significant correlation (.17); therefore, the authors concluded that the magnitude of the VIQ < PIQ discrepancy did not predict or indicate degree of acting out. Andrew (1977) found elevated PIQ, rather than low VIQ, to characterize male delinquents and concluded that any PIQ-VIQ discrepancy could be a source of stress leading to delinquency, regardless of the direction (Haynes & Bensch, 1983)

The pre-1980 research also investigated aspects of neuropsychological functioning in delinquent samples. However, many of the neuropsychological assessments used in these early studies did not include measures of verbal language functioning unless the battery included some version of a Wechsler measure of intelligence (Berman & Siegal, 1976; Hurwitz, Bibace, Wolff, & Rowbotham, 1972). Therefore, although of importance from an historical perspective, the body of pre-1980 research conducted from a neuropsychological standpoint is limited in its contribution to knowledge regarding the contemporary relationship between language functioning and conduct problems.

Despite this limitation, three studies used similar design and procedures to assess the neuropsychological functioning of delinquent adolescents relative to non-delinquent controls

(Berman & Siegal, 1976; Fitzhugh, 1973; Slavin, 1978). All three studies used the Halstead Reitan Neuropsychological Test Battery and some version of the Wechsler scales. Fitzhugh used the Wechsler-Bellevue, whereas the other investigators used the Wechsler Adult Intelligence Scale (WAIS). All three studies found the delinquents consistently demonstrated poorer performance than the controls on verbal and nonverbal Wechsler subtests, as well as poorer performance on the majority of domains assessed by the Halstead Reitan.

Fitzhugh (1973) found deficits in specific areas: speech-sound perception, spatial location, and verbal and nonverbal Wechsler-Bellevue subtests. Berman and Siegal (1976) found that their delinquent group consistently scored significantly below the controls on the WAIS VIQ and PIQ. Also, the delinquent group demonstrated a significant VIQ < PIQ discrepancy, whereas the control group did not. Finally, the delinquent group performed significantly worse than the controls on 6 of 7 measures tapping verbal ability. Slavin (1978), using the same control-group-comparison design, also found the delinquent group to perform more poorly than controls on nearly all WAIS subtests and on 9 of 14 Halstead Reitan tests (Yeudall, Fromm-Auch, & Davies, 1982).

In summary, research conducted prior to 1980 generally supported the conclusion that delinquent youth tend to demonstrate a VIQ < PIQ pattern of intelligence. Studies investigating neuropsychological functioning of delinquents also provided evidence supporting the presence of verbal deficits. This older body of research focused on delinquent samples and primarily used VIQ as an indicator of language functioning. The later research, conducted after 1980, widened the scope of research regarding language functioning in youth with conduct problems. First, researchers began studying youth identified as having CD and ODD, in large part because of the expansion of diagnostic categories for children (e.g., DSM-

III). Second, other standardized methods of assessing language emerged and thus expanded the evaluation possibilities. Language assessment shifted from VIQ as the primary indicator of verbal ability to language batteries addressing global as well as specific aspect of language functioning. Finally, although much of the newer research still includes VIQ, contemporary investigators now can rely on specific neuropsychological measures as well as language-specific measures in their study of the relationship between conduct problems and language integrity.

Review of Post-1980 Research

Language Impairment in Delinquent Youth

As with earlier research, the majority of contemporary investigations have focused on delinquent samples. A summary of these studies is presented in Appendix B. The research reviewed in this section is organized according to the three methods of language assessment: VIQ, neuropsychological measures, and language-specific measures. .

Verbal IQ. As with earlier research, many studies used VIQ as an indicator of verbal ability. A large proportion of this research examined discrepancies between VIQ and PIQ, and its relationship to delinquency. Culberton, Feral, and Gabby (1989) administered the Wechsler Intelligence Scale for Children-Revised (WISC-R) to 82 adolescent boys in a correctional facility. They found 70% of the sample demonstrated a VIQ < PIQ pattern of intelligence; of these, 49% had an 8-point discrepancy, 35% had a 12-point discrepancy, and 26% had a 15-point discrepancy. Similarly, Grace and Sweeney (1986) found 35% of their sample of 20 incarcerated delinquents had a VIQ < PIQ discrepancy of 12 points or more on the WISC-R.

In another study, Cornell and Wilson (1992) administered the WISC-R or WAIS-R to 149 delinquent adolescents, 72 considered violent and 77 considered nonviolent. Of the total sample, 35% obtained a statistically significant VIQ < PIQ discrepancy of at least 12 points. The significance of these findings is highlighted by the fact that a 12 point VIQ < PIQ discrepancy occurred in only 16% of the WISC-R standardization sample. Cornell and Wilson also found that only 5% of the delinquent sample demonstrated the opposite pattern of VIQ > PIQ, compared to 16% of the WISC-R standardization sample.

Wong and Cornell (1999) sought to further characterize the relationship between VIQ < PIQ discrepancy and delinquency by looking specifically at social problem solving and aggression. The sample included 95 male delinquents ranging in age from 13 to 18 years. Measures assessed the participants' intelligence (WISC-R, WISC-III, or WAIS-R), social problem solving, and aggression. Results from inteligence testing showed that 25 % of the sample obtained a significant VIQ < PIQ discrepancy of at least 12 points. Researchers determined that this pattern also related to social problem solving skills (i.e., greater hostile attributional bias), but not to measures of aggression. Walsh, Petee, and Beyer (1987) divided a sample of 256 delinquents into three groups based on a 9-point discrepancy (VIQ < PIQ, VIQ > PIQ, VIQ = PIQ) and compared them on a measure of violence. They found both discrepant groups (VIQ < PIQ and VIQ > PIQ) scored significantly higher on the violence measure than delinquents with no discrepancy.

Another line of research examined the VIQ-PIQ discrepancy in relation to other variables associated with delinquency such as recidivism. Haynes and Bensch (1981) administered the WISC-R to 90 white, male delinquents (36 one-time offenders and 54 recidivists). They found that a VIQ < PIQ discrepancy of at least 15 points occurred more

frequently among the recidivists (70%) than the one-time offenders (42%). Haynes and Bensch (1983) replicated this study using female participants (35 recidivist and 43 non-recidivists). Comparison of these two groups of female offenders indicated that 83% of the recidivists demonstrated a VIQ < PIQ discrepancy versus 58% of the non-recidivists.

Other differences in delinquent behavior, such as degree of violence, have been associated with a VIQ-PIQ discrepancy. Walsh and Beyer (1986) studied 131 juvenile delinquents in terms of their WISC-R scores and degree of violence. Results suggested delinquents with a VIQ < PIQ of 15 points or more (versus delinquents with a VIQ < PIQ discrepancy of less than 15 points) engaged in significantly more delinquent and antisocial behavior, and such behavior started at a significantly earlier age. Petee and Walsh (1987) reached a similar conclusion after assessing the relationship between VIQ-PIQ discrepancy and violent behavior in 125 juvenile delinquents. They used a median split to divide the sample into a high VIQ < PIQ discrepancy group (n = 57) and a low VIQ < PIQ discrepancy group (n = 68). Delinquents identified as having a high discrepancy scored twice as high on a measure of violence than the delinquents in the low discrepancy group.

Findings from other studies examining VIQ supported a link between delinquency and poor verbal ability. Moffitt, Gabrielli, Mednick, and Schulsinger (1981) examined WISC R scores and criminal records of 129 youth from a Danish birth cohort. They found a negative correlation between amount of delinquent involvement and VIQ and Full Scale IQ (FSIQ), but not PIQ, suggesting a specific link with VIQ than overall cognitive ability. These effects remained even after controlling for Social Economic Status (SES). Lynam, Moffitt, & Stouthamer-Loeber (1993) reached a similar conclusion; delinquency was more strongly linked to VIQ than to PIQ in white participants. Denno (1986) examined the records

of 60 African American males convicted of at least one violent offense. A comparison of one-time offenders and repeat offenders revealed that low FSIQ and low VIQ (WISC was administered prospectively at age 7) were the strongest predictors of repeat aggressive offenses against others. Denno also found repeat offenders demonstrated below average achievement and language scores during adolescence. Denno concluded that verbal deficits at an earlier age might be related to the occurrence of both more violent behavior and other criminal behavior.

Dishion, Loeber, Stouthamer-Loeber, and Patterson (1984) investigated the relationship between various skills (e.g., interpersonal problem solving, academic competence, reading, verbal intelligence, homework completion, and chores at home) and delinquency. The sample consisted of 70 tenth-grade boys, 23 delinquents and 47 non-delinquents. Verbal intelligence was measured using the Ammons Full-Range Picture Vocabulary Test. The delinquent group scored lower on six out of the seven skill areas assessed. Deficits in verbal intelligence were significantly correlated with delinquency.

However, results from some studies do not provide supportfor a relationship between VIQ and pattern of juvenile offending. Tarter, Hegedus, Winsten, and Alterman (1985) found no relation between VIQ < PIQ discrepancy and tendency towards violence in their sample of 101 adolescent delinquents. When Cornell and Wilson (1992) compared violent and nonviolent adolescent delinquents in terms of VIQ-PIQ discrepancy, the two groups did not differ in incidence of a VIQ < PIQ pattern of discrepancy. Hubble and Groff (1982) examined the WISC-R scores of 55 male delinquents to determine if any pattern of discrepancy would differentiate between delinquent subgroups as classified by Quay (1979); psychopathic, neurotic, or subcultural. Results indicated that neither magnitude nor

frequency of VIQ-PIQ discrepancy differentiated among the three groups; however, those with psychopathic and subcultural delinquent adjustment subtypes reliably demonstrated higher PIQ than VIQ. It is important to note that, although these studies failed to reveal a connection between VIQ-PIQ discrepancy and specific patterns of delinquent offending, these studies still support a link between VIQ and delinquency.

The inconsistent findings regarding the VIQ < PIQ discrepancy may be attributed to the different methods used to compare groups. Some studies split their samples at the median of the VIQ-PIQ discrepancy scores and then compared the two groups on various indicators of delinquency or antisocial behavior. This places the two groups relatively close together from a data analysis perspective, thus lessening the chance of finding group differences. Also, studies differed in how they define a significant VIQ-PIQ discrepancy. Other studies divided groups based on a statistically significant 12-point discrepancy. Although this number is statistically significant, a 12-point discrepancy is not clinically significant as it occurs with relative frequency in the standardization sample. The frequency of the occurrence of a given discrepancy is often the better indicator of clinical significance.

Making sense of these conflicting results certainly presents a challenge. Culberton, Feral, and Gabby (1989) proposed a few general conclusions after reviewing this literature:

> …the WISC-R is an acceptable diagnostic measure of V/P abilities; the [VIQ < PIQ] difference is not always significant, however, its frequency among delinquents is extensive; the mean VIQ compared with the standardization samples varies by 10-12 points; the [VIQ < PIQ] has not proven to be diagnostic in magnitude in all cases and ranges from 5.6 to 15 points; aggressive and psychopathic subjects show a larger PIQ than VIQ (6 to 20 points); there is growing evidence that the verbal deficits found in delinquents appears to be independent of social class, race, and detection by police (p. 653).

"The robustness of delinquents' deficient VIQs (especially relative to their near-normal PIQs) has been taken as strongly supporting a specific deficit in language manipulation" (Lynam & Henry, 2001, p. 237). However, Seashore's (1951) admonition regarding interpreting VIQ-PIQ discrepancies continues to hold relevance today:

> …we must be extremely cautious in attaching any unusual meaning to difference in Verbal and Performance IQs, even when they are of considerable size. A difference may be important, but not just because it is a difference. Other data must be adduced to permit attaching any import to the discrepancy (either of P > V or V > P) even as big as 5, 10, or 15 points (p. 65).

Neuropsychological measures. Researchers also have examined the language skills of delinquents within the larger context of a neuropsychological assessment. In 1981, Voorhees conducted a study to compare the neuropsychological functioning of 28 juvenile delinquents and a control group of 13 high school students. The participants (boys and girls ranging in age from 13 to 18) underwent a Lurian Neuropsychological Investigation that evaluated functioning in a variety of areas including motor, cutaneous, kinesthetic, visual, receptive speech, expressive speech, reading, writing, arithmetic, memory, and overall intelligence. Results regarding function in 9 of these 11 categories were found to successfully differentiate the delinquent group from the control group. The only two neuropsychological areas that did not differentiate between groups were scores on the cutaneous and kinesthetic domains. The delinquent group exhibited significantly lower receptive and expressive language abilities with specific problems in functions related to verbal integration, word and sentence synthesis, complex grammar, word recognition, and pronunciation of unfamiliar words.

Yeudall, Fromm-Auch, & Davies (1982) also explored possible neuropsychological impairment in juvenile delinquents compared to non-delinquents. Their sample included 99

adolescent delinquent boys and girls (age 13 to 17) in a residential treatment facility for persistent offenders. The control group included 47 adolescent boys and girls from regular education classrooms. Researchers compared the two groups on a number of variables derived from several measures: Halstead Reitan Neuropsychological Battery, WISC-R or WAIS, and 12 other neuropsychological tasks. These measures yielded three language variables pertinent to the current study: VIQ, Word Naming and Memory, and Verbal Fluency and Learning. The delinquent group had a greater percentage with an abnormal neuropsychological profile, but demonstrated a pattern of deficits suggestive of right frontal dysfunction; however, both groups had a VIQ < PIQ pattern.

Wolff, Waber, Bauermeister, Cohen, and Ferber (1982) compared the neuropsychological functioning of delinquent boys and two non-delinquent control groups. The delinquent group consisted of 56 delinquent white males from age 14 to 16 years. The two control groups, a lower-middle class group (n = 48) and an upper-middle class group (n = 48) were matched with the delinquent group on age, sex, and race. The neuropsychological test battery included the following language measures: Boston Naming test, Peabody Picture Vocabulary Test (PPVT), and the Token Test. The delinquent group differed significantly from both control groups on all language measures. Not only did this finding provide further support for language impairment being an important characteristic of delinquents, but it also ruled-out SES as a confounding variable.

Karniski, Levine, Clarke, Palfrey, and Meltzer (1982) administered a neurodevelopmental exam to 54 delinquents and 51 non-delinquent controls. The sample comprised all white males, ranging in age from 12 to 16 years. The neurodevelopmental exam consisted of 29 specific tasks that were used to assess six domains of functioning:

neuro-maturation, gross motor, fine-motor, temporal-sequential organization, visual processing, and auditory-language. The language measures provided indicators of global, receptive, and expressive language ability. Results indicated that the delinquent group performed significantly worse than the control group in the areas of visual processing and auditory-language functioning. However, the greatest difference between the delinquent and control groups occurred in the area of auditory-language functioning. A significantly larger percentage of the delinquent group scored two standard deviations (SD) or more below the comparison group's mean when compared to the control group mean (29.6% versus 2%, respectively).

Robbins, Beck, Pries, Jacobs, and Smith (1983) examined the functioning of 50 adjudicated, non-incarcerated delinquent boys ranging in age from 14 to 18 years across various domains of function (physical, psychiatric, intellectual, academic, perceptual, and motor). One-half of these boys were referred to a mental health clinic for neuropsychological evaluation. Many of the delinquents demonstrated neurological impairment with significant deficits in some aspects of auditory perception (echoic memory, sound discrimination, and speech recognition), which are necessary for the development of verbal skills. Results did not provide evidence for greater impairment among the clinic-referred delinquents compared to those not referred.

Zincus and Gottlieb (1983) investigated the frequency of auditory processing deficits and articulation disorders in a sample of 30 delinquent, institutionalized boys aged 13 to 18. The evaluation consisted of the following measures: WISC-R or WAIS, Wide Range Achievement Test (WRAT), PPVT, Templin Darley Tests of Articulation, and DTLA. Results indicated frequent and significant auditory processing deficits that also related

significantly to academic underachievement. Articulation disorders were found in over 60% of the sample, a striking finding in an adolescent age group.

Tarter, Hegedus, Alterman, and Katz-Garris (1983) examined the neuropsychological functioning of juvenile delinquents with respect to the type of offense committed: violent, nonviolent, or sexual. The sample consisted of 73 male delinquents referred by the courts for a neuropsychological evaluation. Results suggested no difference between these three groups of juvenile offenders on the neuropsychological variables.

Tarter et al. (1984) also explored possible neuropsychological differences between abused adolescent delinquents (n = 27) and non-abused adolescent delinquents (n = 74). The assessment battery consisted of multiple measures tapping cognitive ability (WISC-R), WAIS, Detroit Test of Learning Aptitude-DTLA), achievement (Peabody Individual Achievement Tests-PIAT), impulsivity (Matching Familiar Figures Test-MFFT), neuropsychological functioning (Pittsburgh Initial Neuropsychological Test System-PINTS), and indicators of behavior (Minnesota Multiphasic Personality Inventory-MMPI; Devereux Adolescent Behavior Scale-DABS). Findings showed that the abused delinquents performed significantly worse than the non-abused delinquents with regard to verbal/linguistic processes as indicated by their inferior test scores on VIQ, DTLA, PIAT reading, and PINTS.

Brickman, McManus, Grapentine, and Alessi (1984) administered the Luria Nebraska Neuropsychological Battery to 71 adolescent boys and girls, ranging in age from 14 to 18 years, in a residential setting (64 were included in the analysis). The delinquents demonstrated a pattern of neuropsychological functioning characterized by impaired expressive speech and memory, especially in violent and repeat offenders.

Teichner, Golden, Crum, Azrin, Donohue, and van Hasselt (2000) attempted to identify neuropsychological subtypes of delinquency. They administered the Luria Nebraska Neuropsychological Battery-III to 77 delinquents (mean age of 15.3 years). Of the sample, 77% carried a DSM-IV diagnosis of CD and 17% with ODD. Cluster analysis procedures yielded four neuropsychological clusters: verbal/left hemisphere deficits, subcortical-frontal deficits, mild-verbal deficits, and normal.

Finally, a recent study by Raine, Moffitt, Caspi, Loeber, Stouthamer-Loeber, and Lynam (2005) administered a battery of neurocognitive measures to delinquent boys, with the specific aim of identifying neurocognitive characteristic of different patterns of offending. Participants included 325 adolescent males (mean age of 16 years) from a population-based longitudinal study. Data regarding antisocial behavior, collected from age 7 to 17, was cluster analyzed and resulted in four groups: control (n = 156), childhood-limited (n = 57), adolescent-limited (n = 68), and life-course persistent (n = 44). Then, at age 17 years, participants were administered a battery of neurocognitive measures that included the following: Continuous Performance Task, Wisconsin Card Sort Task, Verbal Dichotic Listening, selected subtests from the Wechsler Memory Scale, and selected subtests from the WISC III. Results indicated that the delinquent youth in this sample demonstrated deficits across neurocognitive domains, with impairments evident in spatial skill, memory, and verbal functioning. In addition, the Life-Course Persistent group of offenders exhibited the most pronounced impairments regarding neurocognitive functioning. Furthermore, the authors concluded that such neurocognitive impairments were not attributable to other factors such as comorbid ADHD, child abuse, psychosocial adversity, or head injury. These findings

provide further support for the differentiation between Adolescent-Limited and Life-Course Persistent delinquency.

Research from the Dunedin, New Zealand Birth Cohort. The Dunedin Multidisciplinary Health and Development Study conducted in New Zealand has contributed tremendously to our understanding of delinquency. This comprehensive longitudinal study began with a 1972-1973 birth cohort of 1036 children in Dunedin, New Zealand. The children were assessed at birth, age 3, and then every 2 years through age 18 years with follow-up rates ranging from 82% to 96% (Silva, 1990). This research was conducted by a group of 44 principal investigators, each being responsible for certain domains of study during the different phases. Domains under investigation include: background and development, physical health (e.g., injury, nutrition, vision, hearing, motor, medical problems, blood analysis, dental, etc.), education (e.g., attainment, academic skills), environmental variables (e.g., SES, school variables, parenting, family relations), risk behavior (e.g., tobacco, alcohol, driving behavior, aggression, delinquency), psychological variables (e.g., attachment objects, self-perceptions, personality, psychiatric status, neuropsychological and cognitive functioning), and others (Silva, 1990).

From the Dunedin data, Moffitt and her colleagues conducted several studies investigating the neuropsychological status of delinquent youth (Moffitt, 1988; Moffitt, Lynam, & Silva, 1994; Moffitt & Silva, 1988a, 1988b, 1988c). These researchers utilized the Dunedin data in different configurations to examine the relationship between neuropsychological status and self-reported delinquency, differences between delinquent and non-delinquent youth, neuropsychological status at age 13 and later delinquent outcome, IQ

and neuropsychological differences between self-reported delinquents and police detected delinquents, and patterns of cognitive deficits associated with delinquent behavior.

The neuropsychological assessment battery consisted of the following tests: WISC-R, Grooved Pegboard, Mazes, Rey Auditory-Verbal Learning Test, Rey-Osterreith Complex Figures Test, Trail-Making Test/Progressive Figures Test, Verbal Fluency, and The Wisconsin Card Sorting Test (WCST). In order to simplify this data set, and because many of the individual scores co-varied, exploratory and confirmatory factor analytic procedures were conducted. This yielded five neuropsychological factors: verbal, visual-spatial, verbal memory, visual-motor integration, and mental flexibility. Most of the relevant studies used the factors, while other studies looked specifically at individual test scores from the battery.

In one study, Moffitt and Silva (1988c) attempted to determine if there was a pattern of cognitive deficits associated with delinquent behavior. They compared four groups across the neuropsychological measures conducted at the age-13 follow-up: No Disorder, ADD-Only, Delinquent-Only, and ADD/Delinquent-Combined. Results indicated that the Delinquent-Only and the ADD/Delinquent-Combined groups scored significantly lower than the No-Disorder control group on 3 of the 5 neuropsychological factors: verbal, visual-spatial, and visual-motor integration. However, the ADD/Delinquent-Combined Group demonstrated the greatest degree of cognitive impairment. Moffitt (1990a) also found that the ADD/Delinquent-Combined Group performed the worst on measures of family adversity, reading, and verbal intelligence, while the Delinquent-Only Group did not demonstrate significant deficits in these areas. These results suggested that the main ingredient in the verbal deficit-delinquency relationship was the presence of ADD.

Moffitt and Silva (1988b) utilized Dunedin data to test the differential detection hypothesis. This hypothesis posits that delinquents who have cognitive or neuropsychological deficits are simply more likely to get caught. Studies using only delinquents that are identified by official police records may introduce sampling bias that can lead to inaccurate conclusions. As already noted, criminal police records do not accurately represent the actual number of crimes committed and it is entirely possible that differences exist between the delinquents who get caught versus those who do not. It may be inaccurate to conclude that verbal deficits are characteristic of delinquents, when such a generalized statement is based on research that employs an unrepresentative sample of the population. Moffitt and Silva explored this issue by comparing the WISC-R IQ scores of 40 detected delinquents (police record), 40 undetected delinquents (self-report), and 545 non-delinquents. Results indicated that the VIQ scores of the detected and undetected delinquents did not differ and that both delinquent groups scored significantly lower on VIQ than the non-delinquents. Thus, results did not support the differential detection hypothesis. They did, however, reiterate the common occurrence of verbal deficits in delinquent youth.

In another study using the Dunedin data, Moffitt and Silva (1988a) examined the relationship between neuropsychological deficits and self-reported delinquency. They compared 124 delinquents and 726 non-delinquents on several self-report measures of delinquent behavior and results of the neuropsychological assessment battery at the age-13 follow-up. Results indicated that self-reported delinquent behavior was associated with a pattern of cognitive deficits in verbal, visual-spatial, visual motor integration, and verbal memory functioning (4 of the 5 neuropsychological factors).

Finally, Moffitt, Lynam, and Silva (1994) provided evidence for a prospective link between early neuropsychological status and later delinquency. They examined the relationship between neuropsychological status for males at age 13, and delinquent outcomes at ages 15 and 18. Whether antisocial behavior was measured with self-report, police records, or court reports, the poorer a boy's neuropsychological functioning at age 13, the more likely he was to have committed crimes at ages 15 and 18. They found that 3 of 5 neuropsychological factors (verbal, visual-spatial, and verbal memory) contributed variance to the prediction of later delinquency beyond what was explained by social disadvantage. They also found that the verbal and verbal memory factors, as measured at age 13, predicted early onset offending (age of first police contact and age of first conviction) as well as persistence of offending. In addition, the 12% of boys with high delinquency and poor neuropsychological status at age 13 were subsequently responsible for 46% of the 251 offenses documented by official police record and 59% of the 255 court convictions. Alternatively, neuropsychological status at age 13 was determined to be unrelated to delinquency that began in adolescence. These results suggested that verbal deficits are related to the LCP pattern of delinquency (and not the AL pattern), and that these deficits better predict delinquency than low SES.

This series of research studies on the Dunedin sample has certain advantages over other studies. The sample consisted of a very large number of unselected male and female participants, thus increasing generalizability of findings. The use of self-report and official police/court records to indicate delinquent status eliminated possible confounds regarding source bias. Researchers conducted neuropsychological testing at age 13, presumably at the beginning of their delinquent career therefore eliminating possible confounds that are present

when testing is conducted with older delinquents. In addition, the neuropsychological tests were selected for the purposes of investigating a specific research question, rather than post hoc collection of existing data. Finally, the longitudinal, prospective design allows researchers to answer questions that cannot be addressed using other designs.

Language-specific measures. Many studies further explore the relationship between language problems and delinquency by directly examining specific language skills. Warr-Leeper et al. (1994) investigated the prevalence of language impairment in a sample of 20 delinquent boys, ranging in age from 10 to 13, in residential treatment placements in Ontario. The assessment battery included the TOAL, TLC, and WISC-R. Results indicated that 80% of the boys demonstrated significant language impairments that had not been previously identified. Deficits in receptive language ability were evidenced by difficulties in listening, understanding abstract language concepts, language without contextual support, language requiring rapid processing, and interpretation of multiple meanings. Expressive deficits were demonstrated by the participants' difficulty with the production of complex language structures that expressed time, reason, and conditional relationships.

In a pilot study, Humber and Snow (2001) compared the language abilities of 15 offenders to 15 controls (all males ranging in age from 13 to 21). The offenders performed significantly worse than controls in every language area as measured by the Test of Language Competence-Expanded (TLC-E) and Speed and Capacity of Language Processing (SCOLP). The offenders demonstrated specific difficulties on tasks requiring quick and accurate language comprehension, decoding abstract language, and providing narrative information logically and sequentially.

A sizable proportion of research in this area has attempted to quantify or describe the need for language intervention services among delinquent youth with unidentified language needs. Sanger et al. (2000) conducted an investigation to understand the communication patterns of incarcerated adolescents. Part of this investigation included the administration of the CELF-3 to 78 participants. Overall, 22% of the adolescents performed 1.3 SD below the mean on the composite. Scores in this range indicated a need for language intervention services, yet none of these youth had ever received language services.

In another study, Sanger, Hux, and Belau (1997) looked at the language abilities of 28 delinquent, adolescent girls with no history of receiving special education services. Scores from the TLC-E indicated that 4 of the girls were potential candidates for language services. Sanger, Moore-Brown, Magnuson, and Svoboda (2001) sought to determine the prevalence of language impairment in a group of incarcerated delinquents. Researchers administered the CELF-3 and the Adolescent Word Test to 67 adolescent females (age ranging from 13 to 18). Results indicated a range of impairment in this sample, with 19.4% obtaining CELF-3 and Word Test scores similar to those meeting eligibility criteria for language services. Results from the Word Test revealed that all girls demonstrated difficulty providing synonyms and definitions for target words, of which some were common, everyday words. Davis, Sanger, and Morris-Friehe (1991) compared the language abilities of institutionalized juvenile delinquents (n = 24) and matched, non-delinquent controls (n = 24). The participants consisted of white males ranging in age from 14 to 17. The groups were matched on age and FSIQ. The TOAL-2 and language sample provided specific information regarding global, receptive, and expressive language functions. Compared to the control group, the delinquents performed significantly worse on this measure. Based on these test scores,

37.5% of the delinquent group qualified for language intervention services, compared to only 4% of the control group.

Stattin and Klackenberg-Larsson (1993) conducted a study to investigate the relationship between early language development and later criminality in a Swedish birth cohort of 122 boys. Assessments began at age 3 and continued through age 17. In addition to standardized measures of language and cognitive functioning, the researchers looked at maternal report of their child's language. In general, when compared to mothers of non-offenders, mothers of offenders reported difficulties in understanding their son's speech at ages 4 and 5 years as well as perceiving their son's language as "backward" (p. 376). Results of this study indicated that early language development was negatively correlated with future criminal behavior, such that a history of language problems may be critical to targeting conduct problem related behaviors.

Summary of studies involving delinquent youth. In studying youth with conduct problems, using samples identified as delinquent continues to be a common practice. In the studies reviewed, the majority utilized male or mostly male adolescent samples. Beyond this similarity, the research reflected a wide range of methodology and design. In general, results from most studies reviewed suggested a relationship between language difficulty and delinquency. A large proportion of this research used VIQ as the primary indicator of language ability. In studies investigating IQ discrepancy patterns in delinquent samples, there seemed to be several important findings. First, the majority of delinquent youth demonstrated a VIQ < PIQ pattern of intelligence. Second, 35% of delinquents obtained a statistically significant 12-point VIQ < PIQ discrepancy, a percentage found in three different investigations (Cornell & Wilson, 1992; Culberton, Feral, & Gabby, 1989; Grace &

Sweeney, 1986). Third, others linked this pattern of discrepancy to impaired social problem solving, increased likelihood of recidivism, more severe delinquent behavior, earlier onset of delinquency, and greater degree of violence. Finally, the relationship between verbal deficit and delinquency appeared to be independent of race and SES.

In contrast, not all of the research provided evidence supporting a relationship between VIQ-PIQ discrepancy and delinquency. Specifically, three studies found no relationship between VIQ < PIQ discrepancy and patterns of offending (e.g., degree of violence). As mentioned earlier, these inconsistent findings may be due to methodological differences (e.g., the specific research question under investigation, how the samples were divided, and definition of a significant discrepancy). While these studies were not able to identify a link between VIQ < PIQ discrepancy and a specific pattern of offending, they still provide evidence for a VIQ < PIQ trend among delinquent youth; therefore, it seems that delinquent youth demonstrate deficient verbal abilities relative to their nonverbal abilities.

Studies investigating language within the larger context of neuropsychological assessment reached a similar conclusion. These studies suggested delinquent youth demonstrated deficits in various domains of neuropsychological functioning, particularly language. Deficits were noted in both expressive language and receptive language. The research conducted using the Dunedin cohort also supported a relationship between impaired verbal ability and delinquency. Delinquent youth (when compared to non-delinquent youth) demonstrated greater impairment in verbal abilities as determined by VIQ and the empirically-defined verbal factor from the neuropsychological assessment battery. Early verbal deficits (as determined by VIQ) were predictive of later delinquency. This finding of verbal dysfunction in delinquent youth remained consistent across different indicators of

delinquent behavior (self-report and official police record) and seemed to be independent of SES. The contribution of comorbid ADD in the relationship between verbal deficit and delinquency remains to be determined.

Finally, several studies evaluated language ability in delinquents using measures designed specifically to assess language functioning. Each of the studies reviewed in this area reached the same conclusion: language difficulty is a significant and common problem in delinquent youth. These youth exhibited difficulties with expressive and receptive language, especially with regard to more abstract contexts. Several investigations demonstrated that a considerable proportion of delinquent youth have significant yet previously unidentified language impairment.

Language Impairment in Youth with Behavioral Problems

Given the findings relating language problems to delinquency, other questions relate to whether these findings extend to other categories of behavior problems and whether individuals actually manifest language impairment. In this regard, one strategy has been to examine the presence of behavior problems in samples of children who already have been identified with language impairments. This body of literature consistently demonstrates that behavior problems commonly occurred in children with language impairment (Baker & Cantwell, 1982; Baker, Cantwell, & Mattison, 1980; Beitchman, Hood, Rochon, & Peterson, 1989; Beitchman, Nair, Clegg, Ferguson, & Patel, 1986; Beitchman, Wilson, Johnson, Atkinson, Young, Adlaf, Escobar, & Douglas, 2001; Benasich, Curtiss, & Tallal, 1993; Cantwell & Baker, 1980, 1985; Carson, Klee, Perry, Donaghy, & Muskina, 1997; Carson, Klee, Perry, Muskina, & Donaghy; 1998; Mattison, Cantwell, & Baker, 1980; Silva, Williams, & McGee, 1987; Stevenson, Richman, & Graham, 1985; Stowe, Arnold, & Ortiz,

2000; Tomblin, Buckwalter, & Catts, 2000). Although the current meta-analysis does not include these studies, which investigate behavior problems in youth already identified with language impairments, this line of research can contribute greatly to our understanding of the relationship between language impairment and conduct problems.

Alternatively, other studies investigate potential language difficulties in youth already identified with behavioral problems. These studies will be reviewed in this section; also, summary information for these studies is presented in Appendix B. Two of the studies reviewed used VIQ as the primary index of language functioning. Three studies assessed language functioning using neuropsychological measures and the remaining three studies utilized language specific measures.

Verbal IQ. The first study that used VIQ as the primary indicator of language function was conducted by Stellern, Marlowe, Jacobs, and Cossairt (1985), with the purpose of investigating hemispheric "cognitive mode," emotional disturbance, classroom behavior, and academic achievement in youth labeled BD compared to normal controls. This study included 94 children in grades three through nine (mean age of 10.5 years); 63% of the sample was male, and 87% of the sample was Caucasian. This sample included a behaviorally disordered group (n = 31) selected from residential schools and a non-disordered control (n = 63) selected from public schools. Participants were assessed using various measures including the Your Style of Thinking and Learning (SOLAT), the Walker Problem Behavior Identification Checklist, the Wide Range Achievement Test, the Bender Gestalt Test, and the WISC-R. The SOLAT is a multiple choice measure which provides respondents with three response styles, one reflecting right-hemisphere processing, another reflecting left-hemisphere processing, and another reflecting integrated processing. This

measure purports to assess hemispheric differences with regard to style of information processing, or "cognitive mode." The authors described the right-hemisphere cognitive mode as "specialized for processing stimuli, especially visuospatial, according to simultaneous and holistic patterns and relationships;" whereas, they described the left-hemisphere cognitive mode as "specialized for processing stimuli, especially language, in terms of sequential, temporal, and feature analysis; and to be dominant for speaking, reading, writing, and arithmetic" (p. 113). Results revealed significant differences between these two groups in terms of their performance across all measures, with the BD youth demonstrating significantly lower VIQ, greater behavioral difficulty, and poorer achievement. The BD youth also demonstrated a preference for right-hemisphere processing, whereas, the control group processed information in a more balanced and integrated manner. The authors concluded that youth with BD exhibit stronger skills with regard to right hemisphere information processing (i.e. visuospatial and simultaneous processing), and consequently, a weakness with regard to left-hemisphere information processing (i.e. language-based and sequential processing).

The second study that used VIQ as the primary indicator of language function was conducted by Cook, Greenberg, and Kusche (1994). These authors investigated the relationship between emotional understanding, disruptive behavior, and intellectual functioning in a school-based sample of first and second grade elementary students. In the initial sample of 220 children (mean age = 8.0 years; 121 males and 99 females), 75% of students participated in regular education while the remaining 25% participated in special education, primarily due to problems related to disruptive behavior, ADHD, and learning disability. The children were classified into one of three groups depending on their level of

disruptive behavior as determined by parent ratings on the Child Behavior Checklist, with particular emphasis on the Externalizing Problems Scale and the Aggression subscale. For example, 18% of the sample was classified as "high behavior problem" (i.e. T-scores > 2 SD above the mean on Externalizing Problems or Aggression); 19% of the sample was classified as "moderate behavior problem" (i.e., T-scores between 1 and 2 SD above the mean on these two scales); and 63% of the sample was classified as "low behavior problem" (i.e., T-scores falling within the average range). Participants were administered a short-form of the WISC-R (includes the Vocabulary, Block Design, and Coding subtests) to provide an estimation of cognitive ability. Results indicated that both the "moderate" and " high" groups performed significantly worse than the "low" group with respect to all three subtests administered, including the Vocabulary subtest. Other results indicated that children with moderate to high levels of behavior difficulty demonstrated significant deficits in emotional understanding, particularly poor recognition of emotional responses in self and others, when compared to the children with low levels of behavior difficulty. These results should be interpreted cautiously due to the methods of identifying youth with behavior problems and the procedures used to classify participants into the three groups. For example, subjective ratings provided by a single informant have the potential to introduce significant bias. Additionally, T-scores of 69 and 70 do not likely represent significant differences in observable behavior, yet these two scores fall within the "borderline significant" and "clinically significant" ranges, respectively.

Neuropsychological measures. Cole, Usher, and Cargo (1993) examined the relation between verbal, visuospatial, and executive functioning and risk for behavior disorders in preschool children (mean age of 56.4 months). Researchers recruited 82 preschoolers with

problems of disobedience, aggression, and misbehavior. The sample was divided into risk groups (high, moderate, and low risk) according to parent and teacher ratings. The participants underwent a large battery of cognitive testing including: McCarthy Scales of Children's Abilities, Florida Kindergarten Screening Battery (FKSB), an executive functioning battery, Diagnostic Analysis of Nonverbal Accuracy-Receptive subtest, and the Forbidden Object task. Results suggested that difficulties in verbal and visuospatial dimensions were significantly associated with occurrence of behavior problems. Specifically, impaired verbal abilities contributed to difficulties with emotional labeling whereas executive functioning predicted behavioral control.

Kusche, Cook and Greenberg (1993) compared the neuropsychological and cognitive functioning of children (age 6 to 10 years old) categorized into four groups according to teacher and self-report ratings: Internalizing (n = 24), Externalizing (n = 62), Combined (n = 27), and Controls (n = 172). The investigators compared the groups on various measures tapping intelligence, academic skills, verbal ability, executive functioning, motor functioning, visuospatial abilities, depression, and classroom functioning. Results suggested that the children in the Externalizing group demonstrated little difficulty on the verbal tasks; especially after excluding children with lower FSIQ from the analysis. However, the instruments used in this study to measure verbal ability (e.g. verbal fluency and Visual Aural Digit Span) are better suited to measure verbal executive functioning. While verbal executive functioning is an important part of language functioning, it is important to note that performance on such measures is influenced by two neuropsychological functions, language and executive functioning. In addition, the use of subjective ratings (self and teacher) to

classify participants into the various groups often introduces significant bias, and is not a preferred methodology for assigning participants to groups.

More recently, Oosterlaan, Scheres, and Sergeant (2005) investigated executive functioning in Dutch youth with disruptive behavior disorders, compared to non-disordered controls. The sample consisted of 99 youth, predominantly male (74%), with a mean age of 10.3 years. The behavior disordered group consisted of 61 participants recruited from special schools for youth with disruptive behavior. Based upon parent and teacher ratings on various measures, the behavior disordered participants were placed into the following groups: ADHD only (36%), ODD/CD only (29.5%), and comorbid ADHD and ODD/CD (34.4%). The control consisted of 38 youth from regular schools. Participants were administered a battery of neuropsychological tasks including Verbal Fluency, Controlled Oral Word Association, SOPT (abstract designs), Tower of London, and the WISC-R. Results indicated that youth rated as ADHD, independent of ODD/CD, exhibited deficits in planning and working memory but not on verbal fluency. With regard to the ODD/CD group, results did not reveal deficits on any tasks, including working memory, planning, or verbal fluency. In fact, many youth rated as ODD/CD demonstrated enhanced performance on measures of executive functioning. The comorbid group generally performed better than the ADHD group but worse than the ODD/CD group, leading to the conclusion that executive functioning deficits in the comorbid group are primarily attributable to ADHD rather than ODD/CD. These findings led authors to conclude that executive functioning deficits are unique to ADHD.

Of note, the placement of this study in this section was judged most appropriate given that the disordered youth in this study were selected on the basis of their non-specific disruptive behavior. Although these participants were described in terms of psychiatric

diagnoses (e.g., ADHD, ODD, and CD), it is important to note that these diagnostic groupings were based on results from parent and teacher ratings, rather than a formal diagnostic evaluation by a trained clinician.

Language-specific measures. Camarata, Hughes, and Ruhl (1988) examined language abilities in 38 children identified as SED by their school. These children ranged in age from 8 to 12 years and all received some amount of special education services. Results from the Test of Language Development-Intermediate (TOLD-I) demonstrated that 97% of the children scored as least one SD below the normative mean on one or more of the subtests and 71% of the sample scored at least two SD below the normative mean on one or more of the subtests. Specifically, the children performed significantly worse on syntactic tasks compared to semantic tasks. They also demonstrated significant deficits on the Speaking Composite relative to the Listening Composite. Although this seemed to reflect a particular profile of language deficits, the sample performed below average on all tasks. To rule-out the confounding effect of low intelligence, investigators studied a subsample of 21 children with FSIQ in the average range and found that 20 obtained scores at least 1 SD below the normative mean on 1 or more subtest.

Minuitti (1991) examined the language abilities of children with language disorders and behavior disorders. The sample (n = 80) ranged in age from 6 to 10 years and was comprised of 3 groups: Language Disordered (n = 27), Behavior Disordered (n = 27), and non-disordered control (n = 26). Language was evaluated using the Clinical Evaluation of Language Functions-Revised (CELF-R) and teachers completed the Behavior Rating Scale. Comparison of CELF-R scores indicated that the Language Disordered and the Behavior Disordered groups did not differ; both performed significantly worse than the control group.

Another comparison was made between children classified as "Language-Deficient" and children classified as "Language-Competent." The Language-Deficient group was defined by any CELF-R score (e.g., Total Score, Receptive Cluster, Expressive Cluster) falling below two standard deviations below the mean of the CELF-R standardization sample. The remainder of the sample was classified as "Language-Competent." Of the 27 children in the Behavior Disordered group, 81% were found to have significant language deficiencies (compared to only 23% of the non-disordered control group). According to teacher rating on the Behavior Rating Scale, the Language-Deficient group demonstrated significantly greater behavioral deviance (i.e., truancy, lying, stealing, property destruction, poor self-control) than children in the Language-Competent group.

Mack and Warr-Leeper (1992) examined the language abilities in an inpatient sample of boys referred for chronic and severe behavior problems (n = 20, ranging in age from 9 to 13 years). The language assessment battery included multiple measures: Token Test for Children, DTLA-2, CELF, PPVT-R, TOAL, TLC, TOLD-I, Expressive One-Word Picture Vocabulary Test (EOWPVT, The Word Test, and Story Reformulation. In this sample, 80% displayed significant impairment in language functioning, a rate 10 times greater than that found in the general population. Comparison between boys with behavioral disorders and language impairment, and boys with behavior disorders and no language impairment revealed that no single language domain differentiated the two groups. However, those with language impairment experienced difficulty on tasks tapping abstract multiple meanings, complex linguistic structures, and meta-linguistic knowledge (e.g., understanding ambiguous sentences, understanding metaphoric expressions, making inferences, listening vocabulary,

speaking vocabulary, recreating sentences, speaking grammar, listening grammar, & oral directions).

Summary of studies involving youth with behavioral problems. All of the studies reviewed in this group investigated possible language impairment in younger (13 years old and younger) children with behavioral disorders. The nature of behavioral problems exhibited by the youth in these studies ranged in severity and chronicity. One study examined the language of preschoolers demonstrating early behavior problems, another looked at inpatient samples with chronic and severe behavior problems, and two studies used youth classified as SED by their school system. Results of these studies pointed to the same conclusion; that is, the majority of children identified as behaviorally disordered either had significant language impairment or were significantly at-risk for language impairment, regardless of the sample source or severity of behavior problems. One study did not reach this conclusion (Kusche, Cook, & Greenberg, 1993); however, the measures used were more indicative of executive functioning rather than verbal ability and classification as behavior disordered was solely determined by rating scales.

Language Impairment in Youth with Oppositional Defiant Disorder

Children with ODD represent a third group of youth with conduct problems. Summary information for the studies examining the language functions in children with ODD is presented in Appendix B. As with the research regarding language problems in children identified as having behavior problems, the research on ODD specifically involves children, but there are fewer studies examining ODD at this time than delinquency or SED.

Speltz et al. (1999) examined the extent to which a pattern of VIQ, language, and executive functioning deficits occurred in a group of preschool boys with ODD compared to

a control group. The ODD group consisted of 80 boys (mean age of 57.1 months); 23 carried a single diagnosis of ODD, 45 carried comorbid diagnoses of ODD and Attention Deficit Hyperactivity Disorder (ADHD), and 12 carried comorbid diagnoses of ODD and some other disorder. The control group consisted of 80 boys matched on age, race, family, and SES. The groups were compared on scores from numerous measures of intelligence, vocabulary, pre-reading, executive functioning, visual-motor skills, and behavior. In general, the clinic-referred boys were more likely than the non-referred boys to exhibit a VIQ < PIQ pattern of intelligence. In looking more specifically at the differences among the clinic referred boys (ODD-Only, ODD/ADHD, and ODD/Other), the comorbid ODD/ADHD boys demonstrated lower verbal and executive functioning scores than the boys with ODD-Only

Similarly, Coy, Speltz, DeKlyen, and Jones (2001) explored social and cognitive characteristics of preschool boys with ODD compared to non-problem peers. The ODD group consisted of 88 clinic-referred boys (mean age of 56.9 months) with and without comorbid diagnoses. The control group consisted of 80 boys (mean age of 57.5 months). Researchers collected data regarding their clinical diagnosis, behavior, social cognition, and verbal ability. Verbal ability was measured with the Comprehension and Arithmetic subtests from the WPPSI-R, the EOWPVT, and the PPVT-R. Results suggested that the clinic-referred boys with ODD demonstrated problems with social-information processing in that they were more likely to generate aggressive solutions and to encode social information less accurately. Verbal IQ, expressive vocabulary, and receptive vocabulary were significantly correlated with tendency to generate aggressive solutions.

Summary of Studies Involving Youth with ODD. Unfortunately, the research on this specific population is sparse. The two studies reviewed supported the idea that children with

ODD tended to also have language problems. Both studies used a control group comparison design to investigate differences in language ability between preschool-age boys diagnosed with ODD and non-disordered controls. Similar to many of the findings from the delinquency and SED studies, the first study found a VIQ < PIQ pattern of cognitive ability among boys diagnosed with ODD (with and without comorbid diagnoses). Also, boys with comorbid diagnoses of ODD and ADHD experienced greater difficulty on verbal and executive functioning tasks than boys with ODD-Only. The second study in this group showed individuals with ODD to have a greater likelihood of problems with social information processing, global language, receptive language, and expressive language. Taken together, these two studies provided continued support for the conjecture that children with conduct problems tend to experience difficulty with language as well.

Language Impairment in Youth with Conduct Disorder

Youth formally diagnosed with CD represent the fourth group of youth with conduct problems. This grouping of studies is important in that criteria used to define CD likely align with research standards as per the criteria in the Diagnostic and Statistical Manual (DSM). The studies examining language impairment in CD are presented in Appendix B. Four studies focused on VIQ as the primary indicator of language ability, while others utilized a neuropsychological approach to this question.

Verbal IQ. Lahey, Loeber, Hart et al. (1995) examined factors related to the persistence of CD across 4 years in a prospective study of clinic-referred boys. They found that low VIQ was related to CD at Time 1. They also found that low VIQ was related to the persistence of CD, especially when low VIQ occurred with a parental history of Antisocial Personality Disorder.

Schonfeld et al. (1988) examined the relationship between cognitive functioning and psychiatric disturbance in 17-year-old black males (n = 115) who were part of a birth cohort. Researchers used archival data to determine IQ at ages 4 and 7, and to obtain health and neurological information. Data regarding intelligence and diagnosis were collected at age 17. Results suggested that the relationship between cognitive functioning and psychiatric status was specific to CD. Three factors accounted for CD at age 17: cognitive functioning measured at both 4 and 7 years of age, parent psychopathology, and early aggression. Furthermore, differences related to CD were more pronounced on the VIQ scales versus the PIQ scales.

Other studies focusing on VIQ did not find verbal deficits in samples of youth with CD. Loeber et al. (1995) reported that VIQ failed to contribute significantly in predicting the onset of CD. Instead, the combination of low SES, previous diagnosis of ODD, and parental substance abuse predicted later diagnosis of CD, with low SES as the strongest predictor. Frick et al. (1994) examined a sample in terms of ODD/CD symptomotology and found a negative correlation between ODD/CD and PIQ, but not VIQ.

Neuropsychological measures. Many researchers examined the language issues suspected in CD using a neuropsychological approach to evaluation. Tramontana and Hooper (1987) examined the neuropsychological functioning of adolescents with CD, depression, and Brian Injury (BI) for two reasons: to determine if those with brain injury could be differentiated successfully based on performance on the Luria-Nebraska Neuropsychological Battery and to compare the neuropsychological profiles of the three groups. The sample consisted of 50 adolescent inpatients divided into 3 diagnostic groups: CD (n = 17), depression (n = 17), and BI (n = 15). Results revealed that 18% of adolescents

in the CD group were classified as neuropsychologically impaired, with expressive language functions being of particular concern in this group.

Werry, Elkind, and Reeves (1987) examined impulsivity, arousal, motor performance, activity level, cognition, and behavior in 95 children, ages 5 to 13, with ADHD, CD, ODD, and anxiety disorders. The sample was divided into three diagnostic groups for comparison: ADHD (n = 39), ADHD+CD/ODD (n = 35), and anxiety disorder (n = 21). The number of children with CD-only or ODD-only was too few for these diagnoses to be considered separate groups (most met criteria for both disorders); therefore, children with CD and/or ODD were considered together as one group (ODD/CD). Results indicated that the ADHD-only group and the ADHD+CD/ODD group differed on Verbal IQ, most behavior ratings, and about 1/3 of the test variables; however, few differences remained when age, sex, and VIQ were partialed out of the analysis.

Frost, Moffitt, and McGee (1989) examined neuropsychological correlates of various psychiatric disorders. Their sample consisted of 678 adolescents (age 13) from the Dunedin, New Zealand study. The authors compared children with no disorder (n = 605), ADD (n = 13), CD (n = 17), anxiety disorder (n = 14), depression (n = 10), and multiple disorders (n = 19). They found no group differences between the CD group and the control group in terms of verbal deficit, and only after including adolescents with CD and comorbid diagnoses did the CD group perform worse than the control group. This suggested that neuropsychological deficits were more strongly associated with comorbidity than CD alone.

Linz, Hooper, Hynd, Isaac, and Gibson (1990) conducted a neuropsychological investigation to determine whether youth with CD demonstrated developmental delay or maturational lag regarding behaviors associated with frontal lobe functioning (executive

functions, attention, response inhibition, planning, working memory). Researchers compared 20 adolescents with CD to 20 control children on nine Lurian tasks. Results indicated no difference between the CD group and control on tasks measuring behaviors attributed to the frontal lobes; however, the CD group performed significantly worse than the control group on receptive vocabulary. Despite the lack of confirming evidence for a developmental lag in frontal lobe functioning, results suggested a relationship between behavior problems and receptive language abilities in children with CD.

Aronowitz et al. (1994) conducted a comprehensive investigation exploring psychiatric, medical, and neuropsychological variables in 20 adolescent inpatients with CD and/or ADHD. The exploration included three group comparisons based on diagnoses of CD and ADHD: CD-Only vs. CD/ADHD, CD-Positive vs. CD-Negative, and ADHD-Positive vs. ADHD-Negative. With regard to the first group comparison, the CD/ADHD group evidenced more neurological soft signs and significant impairment on measures of executive functioning compared to the CD-Onlygroup. A second comparison between the CD-Positive and CD-Negative groups suggested deficits in visuoperceptual and visuospatial abilities among youth with CD. A third group comparison, between ADHD-Positive and ADHD-Negative, identified a pattern of deficit similar to that identified in the first comparison between CD-Only and CD/ADHD participants. None of the groups under investigation demonstrated deficient verbal ability as measured by VIQ.

Dery, Toupin, Pauze, Mercier, and Fortin (1999) conducted a study to investigate a possible association between neuropsychological deficits and CD (with and without comorbid ADHD). A second aim of the study was to explore neuropsychological differences between aggressive and non-aggressive adolescents with CD. The CD group included 59

Canadian boys and girls, ranging in age from 13 to 17 years, recruited from residential facilities and schools providing services to youth with CD. The control group included 29 participants matched on age, gender, and SES. All participants were white and French speaking. The groups were compared using multiple measures of executive functioning and language. The CD adolescents demonstrated significantly lower language skills, but performed similarly to the control group on measures of executive functioning. In addition, youth with comorbid CD/ADHD performed similarly to those with CD-Only; both groups scored at least one-half SD below the control group on certain verbal measures. These findings suggested that deficits in language functioning were attributable to the presence of CD, rather than comorbid ADHD. However, including youth with ADHD in the study increased the chance of selecting a sample with greater impairment in terms of the severity of their CD and antisocial behavior.

In one of the only studies to study the relationship between language functions and CD exclusively in females, Giancola and Mezzich (2000) sought to determine whether adolescent girls with CD differed from non-CD controls in terms of language and executive functioning abilities, and to determine if executive functioning mediated the relationship between language abilities and different types of antisocial behavior. The sample of 320 girls, aged 14 to 18 years, was comprised of two groups: CD group (n = 223) and Control group (n = 97). Researchers compared the groups on measures of language (e.g. TLC-E), executive functioning (e.g. various neuropsychological tasks), and antisocial behavior (e.g. self-report and interview). Findings indicated that girls with CD demonstrated significantly poorer language and executive functioning compared to girls without CD, and that executive functioning mediated the relationship between language competence and antisocial behavior.

Finally, a recent study by Golden and Golden (2001) sought to compare early-onset CD adolescents (n = 15), adolescents with left-hemisphere brain injury (n = 12), adolescents with right-hemisphere brain injury (n = 11), and non-disordered controls (n = 15) in terms of their neuropsychological functioning. The sample was primarily male (60%) and primarily Caucasian (77%), with a mean age of 13.2 years. The participants were assessed using some widely known neuropsychological measures such as the Stroop Color-Word task and the Trail Making task, as well as less known tasks including a general intelligence task, a complex auditory comprehension task, and a vocabulary task. In contrast to the first tasks, which have known reliability and validity, little is known about the latter three tasks beyond the author's description. Results indicated significant differences between the four groups on all tasks administered. Adolescents with CD and adolescents with left-hemisphere brain injury performed the worst, with their performance being significantly different from adolescents with right-hemisphere brain injury and the normal controls. The performance of adolescents with right-hemisphere brain injury fell within the middle. The authors concluded that youth with early-onset CD and youth with left-hemisphere brain injury demonstrate similar patterns of deficits, such that youth with CD may benefit from interventions similar to those used for adolescents with left-hemisphere brain injury. The authors also note that many studies do not differentiate between early- and late-onset conduct problems, possibly explaining some findings of nonsignificant or conflicting results in this field of study.

Summary of studies involving youth with CD. Research investigating language impairment in youth with CD was less conclusive than findings from the other three subgroupings of conduct problems. Two studies used VIQ as the primary indicator of language ability and results suggested that low VIQ predicted later onset of CD especially

when low VIQ occurred with parent psychopathology. On the other hand, results from two other studies investigating VIQ presented conflicting results. Instead of linking VIQ to CD, one found an association between CD and low PIQ and the other found CD to be related to SES, early diagnosis of ODD, and parental substance abuse.

The majority of studies involving CD and language ability went beyond VIQ, relying more on information provided by neuropsychological assessments. Results were mixed in this area as well, with some reporting a relationship between language impairment and CD, while others did not. All found evidence of some type of neuropsychological impairment among youth diagnosed with CD; however, findings differed in terms of the specific nature of the impairment; that is, whether it was in language, executive functioning, or visuoperceptual abilities. Some of these studies had samples comprised of different subgroups reflecting various diagnostic combinations (i.e., CD only, ADHD only, comorbid CD and ADHD). These findings were mixed as well. One study attributed language deficits to CD rather than ADHD, another concluded that deficits were mostly related to comorbidity rather CD alone, yet another did not find evidence to support a verbal deficit in CD youth. Specific language specific measures were included in one study, in addition to other neuropsychological measures, and concluded that executive functioning mediated the relationship between language competence and problem behavior.

These mixed results may be attributed to the inclusion of children with comorbid disorders, especially ADHD. This confounded results, making it difficult to conclude anything specific regarding the relationship between language impairment and CD. Some studies took this problem into consideration by separating participants into diagnostically based subgroups. In these cases, youth diagnosed with CD and a comorbid disorder

demonstrated greater deficits that those carrying a single diagnosis of CD. In one study, the inclusion of CD youth with comorbid disorders made the difference between significance and non-significance in terms of degree of impairment. Also, many of these studies used disordered comparison groups (i.e., ADHD, Depression, and Anxiety) rather than a non-disordered control; thus possibly diminishing the extent to which language problems are associated with CD, but addressing the potential lack of specificity for language problems in a CD population.

Methodological Limitations of Research

Nearly 20 years ago, Moffitt and Silva (1988c) reviewed the methodological limitations of 25 studies investigating the relationship between neuropsychological test scores and juvenile delinquency. Methodological problems regarding subject selection, quality of controls, collection of neuropsychological data, and data analysis raised important questions with regard to interpreting the seemingly consistent findings. Most studies used small, highly selected samples such as incarcerated volunteers, hospitalized delinquents, adjudicated recidivistic offenders, or offenders referred for neuropsychological evaluation. This type of selection targeted the most severe cases and excluded milder cases; therefore, limiting the generalizability of findings. Most of the studies used adolescent samples and did not distinguish between those who began offending in their childhood years and those whose offending did not begin until adolescence. "…These two types are not easily discriminable in adolescence. However, these two types have different developmental trajectories and different psychological profiles; neuropsychological deficits are likely more characteristic of the early starters. Lumping these two types of offenders together will make a large difference (between early starters and controls) seem small (the difference between

adolescent offenders and controls" (Lynam & Henry, 2001, p. 251). Use of older samples also presents several problems. One problem relates to the presence of comorbidity and life events given that older adolescents already have a well-established pattern of delinquent behaviors; that is, there is a greater chance of possible confounds such as substance abuse, head injury, truancy, or institutionalization. Consequently, neuropsychological testing that occurs after years of antisocial behavior cannot provide insight into the causal direction of the relationship between neuropsychological status and delinquency. Similarly, samples obtained from psychiatric settings are more likely to have other concurrent diagnoses (e.g., ADHD, depression) and are also more likely to be taking psychotropic medication. All of these variables have the potential to negatively or positively affect an individual's performance on neuropsychological tests.

Another problem in many studies concerns the quality of the comparison group (Lynam & Henry, 2001). Non-delinquent control groups frequently consist of non-randomly selected volunteers, from higher SES strata (e.g., they have better education). Researchers often use matching procedures to eliminate differences between groups; however, many of the studies did not matched groups on important variables known to influence performance on neuropsychological tests. Alternatively, many studies matched the groups on irrelevant variables. Some studies compared the performance of conduct problem group to test norms of the standardization samples. This type of comparison may not be appropriate if the sample under investigation is not represented within the standardization sample of the test used. Results also depend on the type of comparison group. A non-disordered control group allows for a cleaner, less confounded comparison, and can increase the sensitivity and specificity of the findings.

A third shortcoming of many studies involves the method of collecting neuropsychological information. Many researchers used neuropsychological data that already existed and were collected post hoc from earlier evaluations. This limits what researchers can investigate because they have to use what is there, rather than selecting a battery of tests based on the constructs they want to measure and/or the specific hypothesis that may need to be tested. Cross-sectional studies begin to address this problem, but still do not allow for conclusions regarding the efficacy of neuropsychological test scores as predictors of subsequent conduct problem outcomes.

Study limitations also arise regarding the data analysis procedures. Many studies described the administration of a large battery of tests to a small number of participants and then made group comparisons (use of t-tests) with regard to group performance on each individual test within the larger battery. The use of small samples reduces the power of individual comparisons and increases the likelihood of making a Type II error (accept the null hypothesis of no difference when in fact there is a difference). Alternatively, conducting several individual t-tests increases the likelihood of making a Type I error (rejecting the null hypothesis when there is no difference).

A large liability in the strength of research findings involves the failure to distinguish between conduct problem youth with and without comorbid ADHD. "It may be that deficits reportedly linked to [ADHD] and CP are actually linked to the comorbid group. Because the comorbid group is clearly the most antisocial, failure to consider the comorbidity of [ADHD] and CP actually underestimates the true relation between neuropsychological problems and serious antisocial behavior" (Lynam and Henry, 2001, p. 252). Also, some research findings are likely confounded by medication effects as many youth with ADHD take stimulant

medication, a common treatment for managing inattentive, hyperactive, and impulsive symptoms. Stimulant medication could enhance performance on various tests, especially those measuring attention or processing speed. Indeed, some studies assess for ADHD but do not consider the potential effects of stimulant medication on participant test performance (Giancola & Mezzich, 2000; Raine et al., 2005; Speltz et al., 1999).

Finally, despite the availability of a burgeoning literature over the past 25 years, no meta-analysis regarding language impairment and conduct problems has been conducted to date. Given the wealth of research conducted in this vein, this is a necessary and logical next step. Although it seems that research supports the notion that youth with conduct problems also tend to have problems with language, a quantitative examination of this literature would provide further evidence for this relationship and, perhaps, provide guidance for future research questions in this area of inquiry.

CHAPTER III

METHODS

Meta-Analysis

Meta-analysis is a quantitative method of research synthesis in which individual studies, rather than people, are the "subjects" of analysis. Meta-analysis involves a sequence of stages in the research process that are very similar to original research. For example, meta-analytic stages include problem formulation, collection of data, coding of data, and statistical analysis.

Meta-analysis procedures can only be applied to like findings; that is, findings that are conceptually and statistically comparable (Lipsey & Wilson, 2001). Meta-analysis produces findings in the form of an effect size statistic. There are many different types of effect size statistics; the specific statistic used in a given analysis is determined by the type of finding presented in original research (e.g., correlation, group difference, pre-post difference, and others). The current investigation utilizes the standardized mean difference effect size statistic to analyze findings from studies that compare the language functioning of youth with conduct problems to youth without conduct problems.

Standardized Mean Difference Effect Size Statistic

Effect sizes describing the difference between two groups on some outcome measure generally employ a standardized mean difference effect size statistic from the *d* family. Effect size statistics belonging to the *d* family are computed by dividing the raw difference between group means by an estimate of population standard deviation, which standardizes

the raw difference in group means. Standardization makes it possible to compare studies that operationalize the dependent variable differently (Cooper & Hedges, 1994; Lipsey & Wilson, 2001). In the current study for example, the standardized mean difference effect size enables analysis of studies using various measures of language function in which findings may be standard scores, scaled scores, T-scores, number correct, etc.

The *d* family includes three effect size indices of group difference: Glass's *delta*, Cohen's *d*, and Hedge's *g*. Glass's *delta* standardizes the mean group difference by dividing it by the standard deviation of the population control group. Glass's *delta* is mainly used in meta-analyses of treatment efficacy because of the potential influence of the treatment on the standard deviation of the outcome measure; the standard deviation of the control group remains unchanged (Cooper & Hedges, 1994; Lipsey & Wilson, 2001). Cohen's *d* divides the difference in group means by the pooled standard deviation for the two groups. Cohen's *d* is preferred when study methods are not suspected to significantly alter the outcome measure variance in the experimental group because the pooled standard deviation provides a better estimate of population standard deviation (Cooper & Hedges, 1994; Lipsey & Wilson, 2001). However, Cohen's *d* results in an upwardly biased effect size when calculated for small sample sizes with fewer than 20 participants (Cooper & Hedges, 1994; Lipsey & Wilson, 2001). Hedges's *g,* the effect size statistic used in the current analyses, corrects Cohen's *d* for bias due to sample size.

The standardized mean difference effect size provides an indication of the direction and magnitude of a given research finding. With regard to direction, an effect size can favor the experimental or control group. With regard to magnitude, effect sizes are frequently described according to their range. For example, effect sizes less than .20 indicate little to no

effect, effect sizes between .20 and .49 indicate a small effect, effect sizes between .50 and .79 indicate a moderate effect, and effect sizes greater than .80 indicate a strong effect.

Collection of Data

Literature Search Procedure

The search for potential studies to include in the meta-analysis began with a traditional keyword search using the electronic database PsychINFO. The search was limited to studies published in the English language between 1980 and the present. The search was restricted to this time frame because the third edition of the DSM, published in 1980, was the first edition to include childhood psychiatric diagnoses. Searches were conducted using the following keywords: aggression, behavior, behavior problem, Conduct Disorder, CD, conduct problem, delinquency, disruptive behavior, Emotional-Behavioral Disturbance, EBD, language, neuropsychology, Oppositional Defiant Disorder, ODD, Serious Emotional Disturbance, SED, verbal, and violence. For keywords with multiple derivations, only the root of the word followed by an asterisk was entered. For example, a search for studies with findings on aggression was conducted using the keyword "aggress*" to ensure the search identified articles using the word "aggression" as well as "aggressive." These searches yielded hundreds of references, of which, 107 were deemed potentially relevant.

After collecting these initial references, each article was reviewed for additional references not already obtained. Every potentially relevant reference was recorded, as determined by information cited in the text or by the study title in the reference section. Then, articles were collected, reviewed for their relevance, examined for any additional relevant reference not already obtained, and these ascertained for potential inclusion. This process continued until new references were no longer encountered, resulting in 129

additional potentially relevant references. Ultimately, this literature search process identified a total of 236 articles related to the subject area. Only one of these articles was irretrievable, but this article was identified in the reference list of another article. The resulting pool of 235 articles was then subjected to criteria designed to determine the final pool of articles for inclusion in the meta-analysis procedures.

Inclusion Criteria

Studies were selected for inclusion on the basis of the following inclusion criteria:

1. Study participants 21-years or younger.
2. Studies adhering to a group contrast design where youth with conduct problems were compared to a non-disordered control group.
3. Each study assesses at least one of the following three language constructs: global language, receptive language, or expressive language. Pragmatic language was not considered in the present study.
4. Global, receptive, and/or expressive language functioning was assessed using at least one standardized task. Such tasks may be part of cognitive measures that yielded indices of verbal ability (e.g. Wechsler scales), language measures designed specifically for language assessment (e.g., CELF), and neuropsychological measures that assess language function (e.g. Luria-Nebraska). Such tasks did not include subjective measures (e.g. rating scales) or qualitative indicators (e.g. language samples).
5. The study was published in the year 1980 or later. As previously mentioned, the DSM-III was published in 1980. This was the first DSM edition to include

psychiatric diagnoses specific to child and adolescent populations. Prior to its publication, ODD and CD did not exist as formal diagnoses.

6. Inclusion was limited to published studies, allowing the peer review process to serve as a quality control.
7. The study reported sufficient statistical information to compute an effect size representing the relationship between language functioning and CP.

Of the 235 identified articles, 45 did not study the population of interest. Most of these studies investigated conduct problems in youth with language impairment (rather than language impairment in youth with conduct problems) or youth with general psychiatric difficulties. Forty-six articles were non-empirical (i.e., reviews, book chapters, commentaries) and 21 did not measure language functioning. Eliminating these 112 articles from the pool resulted in 123 relevant empirical studies investigating language function (or neuropsychological function) in youth with conduct problems.

Of these 123 studies, 94 did not meet inclusion criteria regarding design. These studies were excluded for reasons such as: use of qualitative methods, correlational design, factor or cluster analysis, and use of a disordered comparison group (rather than a non-disordered control). While statistical methods make it possible to represent correlational findings in terms of *d,* the present study excludes correlational studies because these studies ask a slightly different question than this investigation seeks to answer. Simply stated, correlational studies ask, What is the relationship between variables X and Y ?; whereas, group contrast studies ask, Do identified groups differ with regard to variable X?

Of the remaining 29 studies, one utilized a rating scale to measure language and five provided insufficient reporting of data needed to calculate effect size. Ultimately, 23 studies

meeting criteria for inclusion remained, representing 18.7% of the 123 relevant empirical studies. However, seven of these 23 studies were based on the same sample. Although all seven studies were coded, only one was retained for analysis to prevent interpretation errors related to statistically dependent effect sizes. The included study was selected because it provided the most comprehensive report of variables of interest, compared to the other six. At the end of this process, 17 studies (13.8% of relevant empirical studies) fulfilled all criteria for inclusion. This process of study selection can be seen in Figure 1 and the included studies are presented in Table 1.

Coding of Studies

In the next phase of the study, relevant information was extracted from each study according to guidelines described in the coding manual that was created for this study (Appendix C). In order to prevent potential bias, a research assistant completed coding. For the purposes of training the research assistant in the coding scheme, one randomly selected article was coded jointly. Then, four studies were randomly selected for the purposes of establishing reliability. The principal investigator and the research assistant coded each study independently and then compared the information extracted. For these four studies, inter-coder agreement was 97%. This level of agreement was expected given the objective nature of the information coded. Coding discrepancies, which occurred for only one study, resulted from discrepant information reported in the original study. For example, the text described sample characteristics at the time of sample selection, whereas the table described sample characteristics at the time of assessment. Because studies frequently contend with issues such as attrition, incomplete participant data, dropping cases during analysis, etc., it was decided to code the information most representative of the time of assessment. This

same guideline was applied to subsequent coding, which was completed by the research assistant.

Information was coded at two levels, study-level and effect size-level. Study-level coding extracts descriptive information regarding study characteristics that apply to the study as a whole (e.g., publication year, sample source, sample demographics, and methods). Each study is associated with a single study-level protocol (Appendix D). At the effect size-level, specific information regarding each individual effect size was extracted (e.g., group means and standard deviations for outcome measure, group size, and type of measure). In contrast to study-level coding, each study may be associated with multiple effect size-level protocols (Appendix E) because each outcome measure produces an effect size; therefore, studies with multiple measures have multiple effect sizes. Original studies did not consistently report certain information. For example, only 47% reported SES, 29 % reported information regarding ADHD, and 24% reported the reliability of outcome measure.

Preparation of Data

Calculation of Effect Sizes

Hedges's *g* was calculated for every measure of language functioning using Comprehensive Meta-Analysis Version 2 (CMA-2), a statistical program developed for the specific purpose of conducting meta-analytic research (Borenstein, Hedges. Higgins, & Rothstein, 2000). The most accurate method for obtaining Hedges's *g* is by direct calculation using the means, variances (or standard deviations), and sample sizes for each group reported in the study. Table 2 presents the original data obtained from studies used in the calculation of *g*. Because all included studies reported this information, no estimation procedures for calculating effect size were necessary. This resulted in a total of 35 effect sizes across 17

studies; however, one outlier was identified and eliminated, as described below. Meta-analyses included 34 effect sizes across 17 studies.

Identification of Outliers

Analysis of effect sizes identified one outlier. The expressive language finding from Golden and Golden (2001) resulted in an effect size of 7.3 with a standard error of 1.007, clearly an outlier when compared to the average effect size (.92) and the average standard error (.09). No explanation for this extreme value could be identified in the original study. Therefore, it was eliminated from analysis, leaving 34 individual effect sizes to be included in analysis. Of note, the average effect size and average standard error reported here do not satisfy the criterion for statistical independence because several studies contributed multiple effect size information. They were calculated for the sole purpose of identifying potential outliers. Outlier analysis at this stage ensures the retention of as many effect sizes as possible. For example, the study by Golden and Golden (2001) resulted in two effect sizes, but only one represented an outlier. The other effect size fell within normal range and was retained in the analysis. Had outlier analysis been conducted after creating independent sets of effect sizes, which requires the averaging of multiple effect sizes from a single study, the Golden and Golden (2001) study would have been eliminated completely.

Creating Independent Sets of Effect Size Data

The next step involved the creation of three independent sets of effect size data, one for each language construct of interest: receptive, expressive and global. The receptive language construct is operationalized according to the description provided in the DSM-IV-TR (APA, 2000), which describes receptive language difficulty as "difficulty understanding words, sentences, or specific types of words…inability to understand basic vocabulary or

simple sentences, and deficits in various areas of auditory processing (e.g., discrimination of sounds, association of sounds and symbols, storage, recall, and sequencing)" (p. 62). Measures providing specific indicators of receptive language include, but are not limited to, the following: Aural Comprehension of Words, CELF, Luria-Nebraska, PPVT, SCOLP, TLC, TOAL, Token Test, and TOLD.

The expressive language construct is also operationalized according to the description provided in the DSM-IV-TR (APA, 2000), which describes expressive language difficulty as a "limited range of vocabulary, difficulty acquiring new words, word-finding or vocabulary errors, shortened sentences, simplified grammatical structures, limited varieties of grammatical structures, limited varieties of sentence types, omissions of critical parts of sentences, use of unusual word order, and slow rate of language development" (p.59). Measures providing specific indicators of expressive language include, but are not limited to, the following: CELF, EOWPVT, Halstead-Reitan, Luria-Nebraska, TOAL, and tasks assessing verbal fluency and verbal naming.

Finally, the global language construct is more loosely operationalized. Similar to Nation, Clarke, & Marshal's (2004) use of the term *"broader language,"* global language is used to indicate any language function. In some instances, this represents measures of language or verbal functioning that cannot be classified as either receptive or expressive language; the Verbal IQ is a primary example as it seems to tap both receptive and expressive language functions. In other instances, global language represents a composite of both receptive and expressive language, as indicated by the Adolescent Language Quotient of the TOAL, the Total Language Composite of the CELF, or the Total Language Composite of the TLC. Finally, global language may represent either receptive language or expressive

language (as indicated by the receptive and expressive language measures enumerated previously). Although it may seem misleading to describe receptive or expressive language as global language, language measures are generally highly correlated (Psychological Corporation, 2003; Sattler, 2001). The high correlations between various language measures suggest, to some degree, that these measures assess a common underlying language function.

In creating an independent set of effect sizes, it is important to ensure that no more than one effect size comes from any subject sample. Including multiple measures representing one construct in a given study can result in errors of interpretation due to statistically dependent effect sizes (Cooper & Hedges, 1994; Lipsey & Wilson, 2001). To prevent this problem, multiple effect sizes for a given construct in a single study were averaged to ensure that each study had no more than one effect size for each language construct. Averaged effect sizes allow for maximum retention of original data. A common criticism of meta-analysis involves the combining of dissimilar findings, often referred to as "mixing apples and oranges" (Sharpe, 1997). However, as previously mentioned, measures of language functioning are usually correlated (Psychological Corporation, 2003; Sattler, 2001), therefore, averaging of findings was judged to be justified.

Data Analysis

Assumptions of Analyses

Analysis of effects and their distributions was conducted under the assumptions of a random effects model (versus a fixed effects model). Briefly, as reported by Cooper & Hedges (1994) and Lipsey & Wilson (2001), fixed effects models assume that the variability in a distribution of effect sizes stems from explainable between-study differences and random error that is solely attributed to subject-level sampling error. Fixed effects models are well-

suited to research aiming to make inferences about similar studies. Random effects models assume that the variability in a distribution of effect sizes results from explainable between-study differences, random subject-level sampling error, and an additional source of unknown random error. The random effects model is more appropriate when the research goals include generalization of findings to the population. The random effects model was selected for the current analyses because of the expectation of multiple sources of random error in the study sample, and because generalization is a primary purpose.

Question 1. What is the magnitude of mean effect size for global language?

Prior to data reduction, the global language data set included 34 separate effect sizes across 17 studies. Nine studies included multiple language measures, and therefore, multiple effect sizes. For each of these nine studies, effect sizes were averaged in order to obtain a single effect size per study. Eight studies included only one measure of language function and did not require averaging. Reduction resulted in a final global language data set of 17 statistically independent effect sizes, one for each of the 17 included studies.

In order to determine the overall effect size for global language, the central tendency and variance for the global effect size distribution were analyzed. Each effect size was weighted by its inverse variance weight before calculating the mean effect size. For the random effects model, the unconditional variance (versus the conditional variance for fixed models) is used in the denominator of the inverse variance weight in order to account for additional random error (Cooper & Hedges, 1994). Distributions of effect sizes are described by various indicators including: range, mean, median, quartiles, standard error, variance, and 95% confidence interval. Forest plots, which display results visually, assist interpretation of

effect size and confidence interval calculated for each study. The significance of the mean effect size is indicated by its *p* - value and 95% confidence interval.

Next, to further investigate the effect size distribution for global language functioning, heterogeneity testing determined whether the various effect sizes that are averaged into a mean value estimate the same population effect size. This test detects the presence of variability in the distribution, above and beyond what is expected from sampling error alone (Cooper & Hedges, 1994; Lipsey & Wilson, 2001). The test for heterogeneity is based on the *Q* statistic, which is distributed as a chi-square with *k* - 1 degrees of freedom where *k* is the number of effect sizes. If *Q* exceeds the critical value for a chi-square with *k* – 1 degrees of freedom, then the null hypothesis of homogeneity is rejected. A statistically significant *Q* indicates a heterogeneous distribution, that the variability within the distribution is affected by additional error beyond sampling error, and requires additional analysis to identify moderator variables. A non-significant *Q* indicates a homogeneous distribution (i.e., the variability in the distribution does not exceed what would be expected from sampling error alone). However, the *Q*-test has relatively low statistical power, especially for analyses including few effect sizes, such that a non-significant *Q* may not accurately predict the absence of one or more moderator variables (Cooper & Hedges, 1994; Lipsey & Wilson, 2001

Question 2. What is the relationship between conduct problems and the specific constructs receptive language and expressive language?

The linkage between conduct problems and effect size magnitudes for receptive and expressive language was investigated by conducting separate meta-analyses. The first meta-analysis included effect sizes describing receptive language. The second meta-analysis

included effect sizes describing expressive language. Each analysis resulted in a mean effect size for its respective language domain, thereby determining the magnitude of mean effect size. Given the group of collected studies, it was not possible to conduct a statistical test for potential difference in mean effect size between receptive and expressive language due to the statistical dependence of effect sizes within these distributions.

The receptive language data set included 12 effect sizes before reduction. For two studies including multiple effect sizes for receptive language, a single effect size was obtained by averaging. Five studies included only one effect size for receptive language. Because no outliers were identified at this stage, the final receptive language data set included seven statistically independent effect sizes, one for each of the studies that measured receptive language. This meta-analysis was conducted in a manner equivalent to the meta-analysis of global language as described above under Question 1, whereby the central tendency of, and variability within, the distribution of effect sizes is analyzed. Each study-level effect size was weighted by its inverse, unconditional variance weight prior to calculating the mean effect size. Other statistics used to interpret the mean effect size included: standard error, 95% confidence interval, and *p*-value. Further, analysis included heterogeneity testing using the *Q*-statistic to assess for variability beyond that expected from sampling error.

The expressive language data set included 10 effect sizes before reduction. Two studies included multiple effect sizes; therefore, the effect sizes in these studies were averaged to produce one effect size for each. Six studies included only one effect size for this construct. The final expressive language data set included eight effect sizes, one for each study that measured expressive language. This meta-analysis was conducted in a manner

equivalent to the meta-analyses of global language, as described above under Question 1; and receptive language, as described in the preceding paragraph. Each study-level effect size was weighted by its inverse, unconditional variance weight prior to calculating the mean effect size. Other statistics used to interpret the mean effect size included: standard error, 95% confidence interval, and *p*-value. Further, analysis included heterogeneity testing using the *Q*-statistic to assess for variability beyond that expected from sampling error.

Question 3. Do certain variables moderate the relationship between conduct problems and the global, receptive, and expressive language constructs?

Moderator analyses were conducted according to theoretically informed a priori decisions, regardless of the results of heterogeneity testing, to determine if the magnitude of the mean effect size for global, receptive, and expressive language differs with regard to age, gender, ethnicity, comorbid ADHD, and type of conduct problem. Of note, the small number of included studies limited moderator analyses in two ways. First, hierarchical moderator analysis was not possible as this would have resulted in too few studies at each level of analysis. Second, moderator analysis was only conducted when the number of included effect sizes was sufficient for meaningful interpretation. This precluded moderator analysis for ADHD and type of conduct problem with specific regard to receptive language and expressive language.

Moderator analyses consisted of meta-regression for continuous variables such as mean age, percent male, percent Caucasian, and percent ADHD. The meta-regression method parallels traditional regression analysis, in which one variable is regressed upon another to determine their relationship, if any. Of note, analyses of ADHD as a moderator only included studies that reported this information (n = 5). A fifth potential moderator, type

of conduct problem, was assessed using the group contrast method. This method applies to categorical variables, rather than continuous variables, and parallels the traditional ANOVA method. Studies were grouped according to the labels used to describe their participant samples (e.g., Delinquent, CD, ODD, and SED/BD). A mean effect size was calculated for each group; then, group means were analyzed for significant between-group variability.

Moderator analyses were conducted using a mixed effects model as well as a fixed effects model because both of these approaches are susceptible to erroneous findings (Cooper & Hedges, 1994; Lipsey & Wilson, 2001). Although mixed effects moderator analysis best suits the purposes of this study, this model may lack sufficient statistical power to detect moderators when applied in meta-analyses with few studies. Alternatively, fixed effects moderator analysis has adequate statistical power to detect between-study differences, but this comes at the risk of high rates of Type I errors, which occur because fixed effects analysis attributes any detected variance to the moderator being tested.

Publication Bias

Meta-analyses for global, receptive, and expressive language were assessed for publication bias. Publication bias often occurs in meta-analytic research because of non-random patterns in published literature. Large studies and studies with moderate to large effects are more likely to be published compared to small studies and studies with non-significant or small effects. Publication bias was assessed using three methods: visual inspection of a funnel plot, Rosenthal's fail-safe N, and Duval and Tweedie's Trim and Fill. Use of multiple methods is recommended because the isolated use of a single method can lead to misinterpretation (Cooper & Hedges, 1994; Lipsey & Wilson, 2001; Soeken & Sripusanapan, 2003).

A funnel plot is a graphic display in which an index of study size (e.g., standard error) is plotted along the vertical axis and the effect size is plotted along the horizontal axis (Coper & Hedges, 1994; Lipsey & Wilson, 2001). Studies with large sample sizes, which have less sampling error, appear near the top of the graph and tend to be located close to the mean effect size. Studies with small sample sizes, which have greater sampling error, appear near the bottom of the graph and tend to have a greater dispersion across the range of effect sizes. Because of these characteristics, a non-biased group of studies is expected to take the shape of an inverted funnel. If visual inspection indicates gaps in the distribution of study plots, publication bias is suspected to occur. For example, a gap on the bottom left side of the graph suggests poor representation of small studies with small effects in the published literature. It should be noted that the shape of plotted studies is difficult to discern in analyses including a small number of studies (Cooper & Hedges, 1994; Lipsey & Wilson, 2001). In such instances, the following methods are necessary supplements in the detection of publication bias.

A second index of publication bias is Rosenthal's (1979) fail-safe N (Cooper & Hedges, 1994; Lipsey & Wilson, 2001; Soeken & Sripusanapan, 2003). This method seeks to address a common concern regarding the "file-drawer problem," in which studies with non-significant results do not get submitted/selected for publication. This method computes the number of studies with non-significant results that would be needed to reduce a significant mean effect size to non-significance ($p > .05$). A large number indicates that, even with the inclusion of many missing studies with non-significant results, a significant effect size would remain significant. However, Gleser & Olkin (1996) recommend cautious interpretation of this estimate, as it often differs substantially from estimates resulting from

other methods and because there is no clear guidelines regarding what constitutes a large number (as cited in Soeken & Sripusanapan, 2003).

The third index of publication bias, Duval and Tweedie's (2000) Trim and Fill method, specifically addresses asymmetry in the distribution of effect sizes (as cited in Soeken & Sripusanapan, 2003). When a meta-analysis includes all relevant studies (i.e., published and non-published, significant and non-significant), the funnel plot indicates a symmetrical distribution of studies on both sides of the mean effect. On the other hand, if the plot appears heavier on the right than the left, the meta-analysis may be missing studies with non-significant results. The Trim and Fill procedure first involves "trimming" the asymmetric studies from the heavy side of the graph in order to locate the unbiased effect. Second, the plot is "filled" by re-inserting the trimmed studies along with their imputed counterparts.

CHAPTER IV

RESULTS

Question 1. What is the magnitude of mean effect size for global language in youth with conduct problems compared to youth without conduct problems?

Meta-Analysis of Global Language

Description of Included Studies.

Meta-analysis of the difference in global language functioning in youth with and without conduct problems included 17 effect sizes. These effect sizes were based on findings from 17 studies, with a combined sample size of 4251 participants. Sample characteristics regarding age, gender, and ethnicity are presented in Table 3. Study samples were comprised of delinquents (n = 7), youth described as SED/BD (n = 5), youth with CD (n = 4), youth with ODD (n = 1).

Analysis of Effect Size Distribution.

Effect sizes for global language ranged from .28 to 1.97 with a median of .89, a mean of .91, and a SD of .12. These values suggest a normally shaped distribution. Table 4 provides additional information describing the distribution of effect sizes for global language. The mean effect size was statistically significant, $g = .91$, $p < .0001$, and indicates a large effect for the standardized difference between global language functioning in youth with and without conduct problems. The 95% CI (.67, 1.15) excludes trivial and small effects. Heterogeneity testing produced a significant Q-value of 112.38, $p < .0001$, which exceeds the chi-square value of 23.54 (16 df). This indicates the presence of significant variability in the

effect size distribution, beyond what would be expected from sampling error alone. Study-level and combined statistics are presented in Table 5. These results are also presented for visual inspection in a forest plot and can be seen in Figure 2.

Publication Bias.

Analysis assessed for publication bias. As suspected, the small number of included studies precluded a meaningful interpretation of the funnel plot by visual inspection alone (Figure 3). However, results from other methods do not support the presence of publication bias. The Trim and Fill procedure did not identify any missing studies on the left side of the plot. Furthermore, the fail-safe N method determined that an additional 1495 studies with non-significant results would be needed to nullify the significance of the mean effect size.

Question 2. Will studies examining the linkages between conduct problems and language impairment evidence any relationship with regard to the specific constructs of receptive and expressive language?

Meta-Analysis of Receptive Language

Description of Included Studies.

Analysis of the difference in receptive language functioning between youth with and without conduct problems included seven effect sizes from seven studies. These effect sizes represented a combined sample size of 460 participants. Sample characteristics regarding age, gender, and ethnicity are presented in Table 3

Analysis of Effect Size Distribution.

Effect sizes for receptive language ranged from .41 to 1.59 with a median of .96, a mean of .92, and a SD of .15. These values suggested a normally shaped distribution. Table

4 provides additional information regarding the central tendency and variance associated with the effect size distribution. The mean value, g = .92, p < .0001, indicated a significant and strong effect size for receptive language, with the 95% CI (.62, 1.22) excluding trivial and small effects. Heterogeneity testing resulted in a significant Q-value of 12.43, p = .05, but it did not exceed the chi-square value of 12.59 (6 df). Although these findings were somewhat mixed, this finding was generally suggestive of a heterogeneous effect size distribution (i.e., presence of significant variance beyond what would be expected from sampling error alone). Study-level and combined statistics for receptive language are presented in Table 6. In addition, results are illustrated in a forest plot in Figure 4.

Publication Bias.

Visual inspection of the funnel plot for receptive language in Figure 5 reveals little about potential bias given the small number of studies. Results from Trim and Fill method suggest three studies missing on the left side of the plot in Figure 6. Re-calculation of mean effect size, including the imputed values, resulted in an imputed point estimate of *g* = .68 (95% CI: .36, 1.01). This estimated *g* suggested a moderate effect size with the 95% Confidence Interval including values indicative of small to strong effects. Results of the fail-safe N method suggested that 135 studies with non-significant results would be needed to reduce the effect size (*g* = .92) to a non-significant value.

Meta-Analysis of Expressive Language

Description of Included Studies.

Analysis of the difference in expressive language functioning between youth with and without conduct problems included eight effect sizes from eight studies. These effect sizes

represent a combined sample size of 915 participants. Sample characteristics regarding age, gender, and ethnicity are presented in Table 3.

Analysis of Effect Size Distribution.

Effect sizes for expressive language range from .28 to 1.20 with a median of .63, a mean of .69, and a SD of .11. These values suggest a normally shaped distribution. Table 4 provides additional information regarding the central tendency and variance associated with the effect size distribution. The mean effect size value, $g = .69$, $p < .0001$, indicated a moderate, but significant effect for expressive language, and the 95% CI (.47, .91) does not exclude a small effect, nor does it exclude a large effect. Heterogeneity testing produced a significant Q-value of 15.21, $p < .05$, which exceeds the chi-square value of 14.07 (7 df). This indicated the presence of significant variability in the effect size distribution beyond what would be expected from sampling error alone. Study-level and combined statistics are presented in Table 7. In addition, a forest plot is provided in Figure 7 to aid visual interpretation.

Publication Bias.

Meta-analysis of expressive language includes assessment for publication bias using three methods. As can be seen in Figure 8, visual inspection of the funnel plot has limited value given the small number of studies included. However, the Trim and Fill and fail-safe N methods were not suggestive of significant publication bias. The Trim and Fill procedure suggested one missing study on the left side (Figure 9). Recalculation of effect size, including the imputed value, provided an adjusted g estimate of .64 (95% CI: .42, .86), which continues to place the effect size within the moderate range. The fail-safe N method

estimated that 175 studies with non-significant results would be needed to nullify the mean effect size calculated from included studies.

Question 3. Do certain key variables (e.g., chronological age, gender, ethnicity, comorbid ADHD, and type of conduct problem) moderate the magnitude of effect size for global, receptive, or expressive language?

Age

Global. Fixed and mixed effects meta-regression procedures were used to test for the potential moderating effect of age. The fixed effects model resulted in a regression line (slope = .05, intercept = .17), indicating a significant positive relationship ($p < .0001$) between mean age and effect size for global language with age accounting for 18.8% of the total variance in the effect size distribution. The mixed effects model resulted in a regression line (slope = .04, intercept = .44), indicating a non-significant relationship ($p = .19$) between age and effect size magnitude with age accounting for only 9% of the total variance in the effect size distribution. Regression lines for the fixed effects and mixed effects models are presented in Figures 10 and 11, respectively.

Receptive. Moderator analysis using fixed and mixed effects meta-regression procedures were used to determine whether age moderates the magnitude of effect size for receptive language in youth with and without conduct problems. Fixed effects and mixed effects modeling results in non-significant findings. The fixed effect model resulted in a regression line (slope = .005, intercept = .78), indicating a non-significant relationship ($p = .78$) between age and effect size magnitude for receptive language, with age accounting for only .6% of the variance observed in the distribution of effect sizes. Similarly, the mixed

effects model resulted in a regression line (slope = -.0007, intercept = .92), indicating a non-significant relationship (p = .98), with age accounting for only .008% of the observed variance. Given the agreement of non-significant findings for both models, regression lines for these analyses are not included.

Expressive. Moderator analysis using fixed and mixed effects meta-regression procedures were used to determine whether age moderates the magnitude of effect size for expressive language in youth with and without conduct problems. The fixed effects model produced a regression line (slope = -.02, intercept = .83), indicating a non-significant relationship (p = .32) between age and effect size for expressive language, with age explaining 6.6% of the observed variance. The mixed effects model produced a regression line (slope = -.02, intercept = .81), also indicating a non-significant relationship (p = .41), with age accounting for 7.9% of the variance in the effect size distribution for expressive language. Given the agreement of non-significant findings for both models, regression lines for these analyses are not included.

Gender

Global. Fixed and mixed effects meta-regression procedures were used to test for potential moderating effect of gender (i.e., percentage of males). The fixed effects model resulted in a regression line (slope = .01, intercept = .50), indicating a significant positive relationship (p < .001) between gender and effect size magnitude for global language, with gender accounting for 10.9% of the total variance observed in the effect size distribution. Results of the mixed effects model resulted in a regression line (slope = .003, intercept - .69), indicating a non-significant relationship (p = .44) between gender and the magnitude of effect size with gender accounting for 3.4% of the variance observed in the effect size distribution.

Regression lines for the fixed effects and mixed effects models are presented in Figures 12 and 13, respectively.

Receptive. Moderator analysis using fixed and mixed effects meta-regression procedures were used to determine whether gender moderates the magnitude of effect size for receptive language in youth with and without conduct problems. Fixed effects and mixed effects modeling results in non-significant findings. The fixed effect model resulted in a regression line (slope = -.006, intercept = 1.35), indicating a non-significant relationship (p = .20), with gender explaining 13% of the variance within the distribution. The mixed effects model produced a regression line (slope = -.006, intercept = 1.34), also indicating a non-significant relationship (p = .37), with gender explaining 10.6% of the observed variance. Given the agreement of non-significant findings for both models, regression lines for these analyses are not included.

Expressive. Moderator analysis using fixed and mixed effects meta-regression procedures were used to determine whether gender moderates the magnitude of effect size for expressive language in youth with and without conduct problems. The fixed effects model produced a regression line (slope = .005, intercept = .26), indicating a non-significant relationship (p = .19) between percentage of males and effect size for expressive language, with gender accounting for 11.1% of the variance. The mixed model produced a regression line (slope = .003, intercept = .49), indicating a non-significant relationship (p = .65), with gender accounting for 2.27% of the variance. Given the agreement of non-significant findings for both models, regression lines for these analyses are not included.

Ethnicity

Global. Fixed and mixed effects meta-regression procedures were used to test for potential moderating effect of ethnicity (i.e. percentage of Caucasians). Results of fixed effects and mixed effects moderator analyses were in agreement. The fixed effects model resulted in a regression line (slope = -.01, intercept = 1.72), indicating a significant negative relationship ($p < .0001$) between percentage of Caucasians and effect size for global language functioning, with 31% of the variance in the effect size distribution explained by ethnicity. The mixed effects model produced a regression line (slope = -.01, intercept = 1.62), also indicating a significant negative relationship ($p < .05$), with ethnicity explaining 22.7% of the variance. Regression lines for the fixed effects and mixed effects models are presented in Figures 14 and 15, respectively.

Receptive. Fixed and mixed effects meta-regression procedures were used to determine whether ethnicity moderates the magnitude of effect size for receptive language in youth with and without conduct problems. The models resulted in nearly equivalent findings. The fixed effects model produced a regression line (slope = -.01, intercept = 1.56), indicating a significant negative relationship ($p = .01$), with percentage of Caucasians explaining 50.8% of the variance within the distribution of effect sizes for receptive language. Similarly, the mixed effect model resulted in a regression line (slope = -.01, intercept = 1.58), indicating a significant negative relationship ($p = .01$), with percentage of Caucasians explaining 51.1% of the variance. Regression lines for the fixed and mixed models of ethnicity are presented in Figures 16 and 17, respectively.

Expressive. Moderator analysis using fixed and mixed effects meta-regression procedures were used to determine whether ethnicity moderates the magnitude of effect size for expressive language in youth with and without conduct problems. The fixed effects

model resulted in a regression line (slope = -.004, intercept = .95), indicating a non-significant relationship (p = .22) between percentage of Caucasians and expressive language, with ethnicity explaining 9.94% of the observed variance in the distribution. The mixed model produced a regression line (slope = -.006, intercept = 1.13), also indicating a non-significant relationship (p = .16), with ethnicity explaining 24.2% of the observed variation. Given the agreement of non-significant findings for both models, regression lines for these analyses are not included.

ADHD

Global. Fixed and mixed effects meta-regression procedures were used to test for potential moderating effect of ADHD (i.e., percentage of disordered group with ADHD). This analysis only included studies that reported information regarding this diagnosis (n = 5). A fixed effects model produced a regression line (slope = -.02, intercept = 1.53), suggesting a significant (p < .0001) negative relationship between percentage of ADHD and effect size magnitude, with ADHD contributing to 43.3% of the variance in effect size. Results from the mixed effects model produced a regression line (slope = -.02, intercept = 1.42), which also indicates a significant negative relationship (p = .05), with ADHD contributing 43.1% of the variance in effect size. Regression lines for the fixed effects and mixed effects models are presented in Figures 18 and 19, respectively.

Receptive. Only two studies in the receptive language meta-analysis report information regarding ADHD, therefore, analysis does not address this specific question.

Expressive. Only three studies in the expressive language meta-analysis report information regarding ADHD; therefore, analysis does not address this specific question.

Type of Conduct Problem

Global. Moderator analysis compared studies according to type of conduct problem using a mixed effects group contrast method. Results from this analysis are presented in Table 8. Articles were separated into the following four groups: delinquent (n = 7), SED/BD (n = 5), CD (n = 4), and ODD (n = 1). Results for delinquents revealed a significantly large effect size, g = 1.03, p < .001 (95% CI: .62, 1.44); for SED/BD, the effect size was significant and moderate, g = .77, p < .001 (95% CI: .30, 1.25); for CD, the effect size was significantly large, g = .98, p < .001 (95% CI: .42, 1.54); and for ODD the effect size was not significant, but it was moderate in strength, g = .67, *ns* (95% CI: -.36, 1.70). Of note, the effect size for the ODD group of articles was based on a single study and does not represent a mean effect size. Heterogeneity analysis contrasting the mean effect between these groups results in a Q-value of .93, which does not exceed the chi-square value of 7.82 (df 3). This indicates the absence of significant variability between groups. Therefore, type of conduct problem does not account for the variability observed in the distribution of effect sizes.

Receptive. Given the small number of studies in the receptive language meta-analysis, further division of effect sizes by type of conduct problem results in very small group sizes. Therefore, analysis does not address this specific question.

Expressive. Given the small number of studies in the expressive language meta-analysis, further division of effect sizes by type of conduct problem results in very small group sizes. Therefore, analysis does not address this specific question.

CHAPTER V

DISCUSSION

Meta-analytic procedures were used to synthesize research findings from studies investigating language functioning of youth with conduct problems compared to youth without conduct problems. More specifically, analyses were conducted with the purpose of describing the nature of the following three language constructs: global language, receptive language, and expressive language. Seventeen studies met criteria for inclusion. From these 17 studies, three independent sets of effect sizes were created, one for each language construct. Each mean effect size and its accompanying distribution were then analyzed for heterogeneity, followed by moderator analyses as indicated by a priori hypotheses. Analysis also included testing for publication bias.

Question 1. What is the magnitude of mean effect size for global language in youth with conduct problems compared to youth without conduct problems?

Meta-Analysis of Global Language

Meta-analysis of global language included 17 effect sizes from 17 studies. Results indicated a significant mean effect size of large magnitude for the difference in global language functioning between youth with and without conduct problems (g = .91; 95% CI: .67, 1.15). Results of heterogeneity testing identified significant variability in the magnitude of observed effects and were suggestive of potential moderating variables, with subsequent analyses identifying these moderator variables. Trim and Fill and fail-safe N procedures did not support the presence of publication bias.

It was hypothesized that findings would show a significant mean effect size of at least moderate magnitude for global language impairment between youth with and without conduct problems. The finding of a significant and large mean effect size confirms this hypothesis. In fact, the magnitude of the mean effect size, g = .91, exceeds the hypothesized moderate effect size; although the 95% CI does not exclude a moderate effect. The finding of a significant mean effect size also confirms most previous research investigating the language functioning of youth with conduct problems. More specifically, these results provide strong support for studies concluding that a large proportion of youth with conduct problems have unidentified language impairment, or are at-risk for such language impairment (Davis, Sanger, & Morris-Friehe, 1991; Miniutti, 1991; Sanger et al., 2000, 2001; Sanger, Hux, and Belau, 1997). Overall, the research literature provides strong evidence for the presence of language difficulties in this population. Given the magnitude of demonstrated mean effect size, the inclusion of widely different subject samples, and the nature of the study design, the finding of impaired global language functioning in youth with conduct problems is highly generalizable to the broader population of youth with conduct problems.

The results of this meta-analysis suggest that many youth with conduct problems have a specific underlying neuropsychological deficit (i.e., problems with language processing). Therefore, treatment approaches that address this underlying deficit as part of a larger treatment program may prove more effective. Language deficits should be identified as early as possible in order to maximize treatment progress and long-term outcome. Therefore, all young children with problem behavior should be assessed for language problems. Schools are a logical setting for this assessment as children with problem behavior are generally identified at young ages.

Question 2. Will studies examining the linkages between conduct problems and language impairment evidence any relationship with regard to specific constructs of receptive and expressive language?

Meta-Analysis of Receptive Language

Meta-analysis for receptive language included seven effect sizes from seven studies resulting in a significant mean effect size of strong magnitude (g = .92, p < .0001), and the 95% CI (.62, 1.22) excluding trivial and small effects. Heterogeneity testing indicated significant variability within the effect size distribution, beyond sampling error, which was suggestive of potential moderating variables. Subsequent analyses further investigated the potential presence of such moderators. Possible publication bias was assessed statistically using two methods. The classic fail-safe N method was not suggestive of publication bias. The Trim and Fill method suggested three missing studies, imputed an effect size value for each, and used these imputed values to recalculate the mean effect size. This procedure resulted in an estimated mean effect size of g = .67 (95% CI: .36, 1.01), which is indicative of a moderate effect.

It was hypothesized that meta-analysis would show a significant mean effect size for the receptive language functioning between youth with and without conduct problems. The finding of a significant and strong mean effect size confirms this hypothesis and provides further support for the conclusions of previous studies documenting receptive language impairment in youth with conduct problems (Humber & Snow, 2000; Linz et al., 1990; Voorhees, 1981; Warr-Leeper et al., 1994). While it is possible that publication bias contributed to an inflated mean effect size, the recalculation of mean effect size using imputed values still resulted in an effect size of moderate strength. Additionally, small meta-analyses have low statistical power and are

susceptible to errors of under-detection. Given this, these results provide strong evidence for receptive language difficulties in youth with conduct problems.

Meta-Analysis of Expressive Language

Meta-analysis for expressive language included eight effect sizes from eight studies. Analysis resulted in a significant mean effect size of moderate magnitude (g = .69, p < .0001). The 95% CI (.47, .91) does not exclude a small effect, nor does it exclude a large effect. Heterogeneity testing indicated the presence of significant variability in the effect size distribution, beyond what would be expected from sampling error alone. This was further assessed via moderator analyses. Methods for assessing publication bias were not suggestive of a significant number of missing studies.

It was hypothesized that meta-analysis would show a significant mean effect size for expressive language functioning between youth with and without conduct problems. The resulting significant, moderate mean effect size supports this hypothesis. Given the low statistical power of the meta-analysis for expressive language, findings provide strong support for expressive language deficits in many youth with conduct problems, a conclusion that aligns with previous research findings (Brickman et al., 1994; Camarata, Hughes, & Ruhl, 1988; Tramontana & Hooper, 1987; Voorhees, 1981; Warr-Leeper et al., 1994; Zincus & Gottleib, 1983).

Additionally, it was hypothesized that the mean effect sizes for receptive and expressive language would not differ significantly. It was not possible to establish statistically independent sets of effect sizes for receptive and expressive language from the collected sample of studies, which precluded group contrast of mean effect sizes for determining statistical difference between these language domains. Given this, analyses were not able to confirm or refute the

hypothesized equality in mean effect size for receptive and expressive language. Although it is not possible to provide a definitive conclusion regarding potential differences between domains, the mean effect size values for these constructs, and the little overlap in their 95% confidence intervals, suggest that youth with conduct problems exhibit poorer receptive language compared to expressive language.

Question 3: Do certain key variables (e.g., age, gender, ethnicity, comorbid ADHD, and type of conduct problem) moderate the magnitude of effect size for global, receptive, or expressive language?

Age.

It was suspected that younger participant samples would exhibit stronger effect sizes than older participant samples for each of the three language constructs. Fixed effects and mixed effects meta-regression were conducted for each of the three language constructs, for a total of six analyses. Only one of these, the fixed effects model for global language, resulted in significant findings. Given that fixed models are susceptible to Type I error (i.e., identifying a significant relationship when none exists), these results suggest that age does not moderate the magnitude in effect size for the language functioning between youth with and without conduct problems.

These findings must be interpreted cautiously given that this meta-analysis is more heavily weighted by studies employing adolescent samples (n = 10) compared to child samples (n = 4), with the remaining studies employing samples that include both children and adolescents (n = 3). Furthermore, studies with primarily adolescent samples were some of the largest samples included in this investigation (Moffitt, 1988; Raine, 2005). The greater contribution of adolescent age groups versus child age groups in the current study is an especially important

consideration in light of previous research establishing a link between language deficit and early onset conduct problems, but not late onset conduct problems (Moffitt, Lynam, & Silva, 1994). Therefore, these results should not be interpreted to mean that chronological age does not contribute to the relationship between language functioning and conduct problems.

Although the absence of certain findings provides limited basis from which to draw conclusions, nonsignificant findings should not be disregarded as uninformative. Indeed, both models implicitly agreed that global language functioning in youth with conduct problems does not improve over time. This finding is especially important when considered in conjunction with research demonstrating the persistence of conduct problems (Caspi et al., 1987; Farrington & Loeber, 2000; Loeber, 1982, 1991; Loeber et al., 1995; Tremblay et al., 1991), which also may indicate low treatment effectiveness and/or a treatment recalcitrant condition. Although the efficacy of some interventions is supported by research (e.g., parent training in behavior management and anger coping), these interventions do not produce lasting gains, especially for youth with more severe conduct problems. These conclusions provide a compelling argument favoring a new approach to intervention.

Gender

It was hypothesized that studies with higher percentages of males would show larger effects than studies with lower percentages of males for each of the three language constructs. Moderator analyses using fixed effects and mixed effects meta-regression were conducted for each language construct, for a total of six analyses. Only one of these, the fixed effects model for global language, resulted in significant findings. Given that fixed models are susceptible to Type I error (i.e., identifying a significant relationship when none exists), these results suggest that gender does not moderate the magnitude in effect size for the language functioning between

youth with and without conduct problems. The lack of a significant relationship may suggest that males and females with conduct problems are equally likely to have unidentified language impairment.

It is important to note that females are underrepresented in this literature area, with most previous studies employing all-male samples (Coy et al., 2001; Culberton, Feral, & Gabby, 1989; Davis, Sanger, & Morris-Friehe, 1991; Denno, 1986; Dishion et al., 1984; Haynes & Bensch, 1981; Humber & Snow, 2001; Lahey et al., 1995; Karniski et al., 1982; Robbins et al.,1983; Speltz et al.,1999; Stattin & Klackenberg-Larsson, 1993; Wolff et al.,1982; Warr-Leeper et al., 1994; Wong & Cornell, 1999) and relatively few employing all-female samples (Giancola & Mezzich, 2000; Haynes & Bensch, 1983; Sanger, Hux, Belau, 1997; Sanger et al,, 2001). This imbalance was evident in the studies selected for the present investigation. Given this, the studies included in this meta-analysis may not provide an accurate reflection of females with conduct problems.

Ethnicity

It was hypothesized that studies with lower percentages of Caucasian participants would show larger effects than studies with higher percentages of Caucasian participants for each of the three language constructs. For global and receptive language, the fixed effects and mixed effects meta-regression models indicated a significant negative relationship between effect size and percentage of Caucasian participants (i.e., the magnitude of effect decreases as the proportion of Caucasian participants in the study increases). Results for expressive language were not suggestive of a significant relationship. These findings suggest that minority youth (i.e., of African American or Hispanic descent) with conduct problems demonstrate language difficulties in their global and receptive language functioning but not expressive language functioning.

Of note, minority youth are generally under-represented in the research literature. For example, among the studies reviewed that reported information regarding the ethnicity of participant samples, most utilized Caucasian-Only participant samples (n = 23) or samples with greater than 60% Caucasian participants (n = 17); while relatively few studies utilized samples predominated by African American participants (n = 8) or Hispanic participants (n = 1). Similarly, the studies included in the current meta-analytic investigation reflect this disproportion; over one-half of included studies utilized Caucasian-Only samples (n = 10). Also, only five of the studies included in this meta-analysis reported no significant difference between the disordered group and non-disordered group with respect to ethnicity. Furthermore, SES, which may contribute to both language impairment as well as conduct problems, was not included as a variable in the current analysis. As such, it is possible that low SES explains the observed relationship between language functioning and conduct problems in minority youth. Given the imbalance with regard to under-representation of ethnic minorities, as well as potential confounds such as SES, the finding of ethnicity as a moderating variable between language functioning and effect size should be cautiously interpreted.

ADHD

It was hypothesized that effect size would increase in magnitude as percentage of participants with ADHD in study samples increased. Fixed effects and mixed effects meta-regression analyses suggested a negative relationship between global language and ADHD, with the fixed model indicating a significant relationship and the mixed model indicating a nearly significant relationship. These findings suggestthat samples with larger percentages of youth with ADHD showed smaller effect sizes. Analyses did not investigate the potential moderating

effect of ADHD with regard to receptive language or expressive language due to small number of studies that reported this information.

These results suggest that ADHD has some contribution in the relationship between conduct problems and language impairment. However, the nature of this relationship remains unclear given the mixed findings in this area, with some suggesting a positive relationship (Moffitt, 1990; Moffitt & Silva, 1988c; Speltz et al., 1999) and others suggesting no relationship (Aronowitz et al., 1994; Dery et al., 1999; Frost, Moffitt, & McGee, 1989; Oosterlaan, Scheres, & Sergeant, 2005). Either way, ADHD appears to contribute its own unique variance to group differences such that comorbidity should be controlled in studies examining conduct problem populations and language impairment.

Type of Conduct Problem

It was suspected that meta-analytic findings would show significant differences in mean effect sizes for language impairment across the different types of conduct problems. More specifically, results were expected to show larger mean effect sizes for CD and ODD, compared to mean effect sizes for Delinquent and SED/BD. Moderator analysis for type of conduct problem was conducted only for the global language construct because of the small number of effect sizes included in the receptive and expressive language distributions.

Groupings of studies indicated large mean effect sizes for youth labeled delinquent, CD, and BD. A moderate effect size was found for youth labeled ODD; however, the observed effect for this grouping represented findings from only one study (rather than an average effect for multiple studies). Moderator analysis investigating between-group variability in the observed mean effect sizes for global language indicated no significant difference between youth labeled delinquent, CD, BD, and ODD; therefore, results do not support the significant differences in

mean effect size between these groups as hypothesized. This finding highlights the similarity between youth with differently-named conduct problems regarding language impairment. This is indeed interesting given that the labels derive from such various disciplines.

Future Directions

Research should continue to investigate the language functioning of youth with conduct problems. The current study, along with the many that precede it, provides strong evidence for language difficulties in youth with conduct problems; however, a meta-analysis including both correlational studies and group contrast studies (with findings from both represented as effect size statistic *d*) would further solidify this link given the increased statistical power associated with larger sample size. An investigation of this nature would be helpful in describing the association between language functioning and conduct problems in youth, as well as the different factors that potentially influence this association. Furthermore, a larger meta-analysis would permit hierarchical moderator analysis, a more sophisticated method for identifying moderator variables.

Additional research is also needed to elucidate potential differences with respect to more differentiated language functions. Most broadly, receptive and expressive language constructs require further study. An important question for future research is to determine the specific cognitive correlates for receptive and expressive language. For example, effective expressive language partially depends on executive functioning (i.e., organization, planning, and fluency); whereas, effective receptive language functioning partially relies on the ability to interpret nonverbal communication (i.e., facial expressions, body language, and intonation). This latter example suggests a need for further investigation of pragmatic language functioning in this population, particularly since pragmatic language represents an interaction of language processes

and executive functioning within the context of social interaction. Pragmatic language functioning is a relatively new area of study; therefore, many questions remain regarding its possible role in the development, expression, and maintenance of conduct problems.

It is especially important to examine the mechanisms by which these language functions manifest in youth with conduct problems, how they develop, and how they inter-relate over time, particularly with respect to the emergence of problem behaviors.

Many questions also remain regarding other neuropsychological functions in youth with conduct problems. Executive functioning deficit is a second relevant construct in youth with conduct problems (Aronowitz et al., 1994; Cole, Usher, & Cargo, 1993; Dery et al., 1999; Giancola & Mezzich, 2000; Speltz et al., 1999). Just as meta-analysis was a logical next step in reaching a conclusion regarding the language functioning in youth with conduct problems, meta-analysis regarding executive functioning is not likely far behind. While language and executive functioning appear the most relevant neuropsychological constructs in the study of conduct problems, visual-spatial abilities cannot be excluded (Cole, Usher & Cargo, 1993; Karniski et al., 1982; Moffitt & Silva, 1988a, 1988c; Voorhees, 1981). In order to fully understand the neuropsychological functioning of youth with conduct problems, it is important for research to continue to extend beyond Verbal IQ. In this manner, research can continue to move forward in identifying subtypes of conduct problems. Such research is important in predicting those at greatest risk.

Continued investigation regarding the roles of comorbid disorders is also needed. The specific contribution of ADHD remains in question. It may be wise to consider the different ADHD subtypes (e.g., Inattentive, Hyperactive-Impulsive, and Combined) as part of future research analysis. Another important comorbidity to consider is learning disability, given its

association with conduct problems and its association with language problems. A third comorbidity to investigate is language impairment. While the current study provides evidence for poorer performance of youth with conduct problems on language measures compared to youth without conduct problems, it does not clarify the relationship between actual language impairment (as defined by DSM criteria) and conduct problems. Given the wealth of existing studies investigating the development of conduct problems in youth with language impairment, meta-analytic procedures may be well-applied to this literature as well.

Finally, each passing year results in a multitude of new research findings for a given domain. Therefore, quantitative research synthesis will become increasingly important in the integration of primary research findings. Researchers of primary research can facilitate this integrative process by consistently and clearly reporting data that are commonly of interest to meta-analysts.

Table 1.

Included Studies

Study	Sample Description	Language Measures
Cook et al. (1994)	N = 220 (full data for 213) Disordered: students rated >1 SD above the mean on externalizing behavior Control: students rated < 1 SD above the mean on externalizing behavior Age: mean = 8.0 years Matching: not matched Gender: 55% male Ethnicity: 67.3% Caucasian, 24.5% African American, 8.2% other	WISC R (Vocabulary)
Davis et al. (1991)	N = 48 Disordered: institutionalized delinquents (n = 24) Control: junior and senior high school students (n = 24) Age: mean = 16.6 years Matching: age, FSIQ, gender*, & ethnicity* Gender: 100% male Ethnicity: 100% Caucasian	TOAL-2 (Total Language Quotient, Receptive Language Quotient, & Expressive Language Quotient)

Dery et al. (1999)	N = 88 Disordered: CD youth from various treatment settings (n = 59); ADHD (23.7%) Control: public school students in regular education (n = 29) Age: mean = 15.3; range = 13 to 17 years Matching: age*, gender*, SES*, & ethnicity* Gender: 77% male Ethnicity: 100% Caucasian (French Canadian)	Aural Comprehension of Words, Token Test, Visual Naming, & Controlled Oral Word Association
Giancola & Mezzich (2000)	N = 320 Disordered: CD females from multiple settings (n =22); ADHD (17%), anxiety disorders (37%), Depression (35%), Adjustment Disorder (10%), Dysthymia (9%), eating disorder (8%), Bipolar Disorder (.4%) & substance use disorders (77%) Control: obtained through a recruitment agency (n = 97); Adjustment Disorder (4%) Age: mean = 16.0 years; range = 14 to 18 years Matching: gender* Gender: 100% female Ethnicity: 71% Caucasian & 26% African American	TLC-E (Total Score)

Golden & Golden (2001)	N = 30 Disordered: CD youth from a mental health clinic (n = 15) Control: middle school students in regular education (n = 15) Age: mean = 13.2 years; range = 11 to 14 years Matching: not matched Gender: 60% males Ethnicity: 76.7% Caucasian, 10% African American, 10% Hispanic, & 3% other	vocabulary task & auditory comprehension task
Humber & Snow (2001)	N = 30 Disordered: adjudicated delinquents (n = 15) Control: public school students (n = 15) Age: mean = 16.45 years; range = 13 to 21 years Matching: age*, gender* Gender: 100% male Ethnicity: 100% Caucasian (Australian)	SCOLP

Karniski et al. (1982)	N = 105 Disordered: newly committed delinquents (n = 54) Control: middle and high school students (n = 51) Age: mean = 14.7 years; range = 12 to 16.5 years Matching: gender* & ethnicity* Gender: 100% male Ethnicity: 100% Caucasian	Auditory Language Function Composite (Sentence Repetition, Syntax Comprehension, Token Test, Verbal Opposites, & Boston Naming Test)
Kusche et al. (1993)	N = 281 Disordered: BD students with externalizing problems in special education (n = 109) Control: regular education students (n = 172) Age: mean = 7.9 years Matching: not matched Gender: 60% male Ethnicity: 61% Caucasian, 30% African American, 6% Asian, 3% Native American or Hispanic	WISC R (Vocabulary) & McCarthy Scales (Verbal Fluency)
Linz et al. (1990)	N = 40 Age: mean = 15.6 years; range not reported Disordered: CD youth in an evaluation center (n = 20)	PPVT-R

	Control: youth from a variety of locations (n = 20) Matching: age, race, & gender* Gender: 50% male Ethnicity: 50% Caucasian	
Miniutti (1991)	N = 53 Disordered: BD students in special education (n = 27) Control: students in regular education (n = 26) Age: mean = 7.7 years; range = 6 to 9 years Matching: age* Gender: 66% male Ethnicity: 76.9% African American, 13.0% Caucasian, & 10.0% Hispanic	CELF-R (Total Composite, Receptive Composite, & Expressive Composite)
Moffitt & Silva (1988b)	N = 654 Disordered: delinquents from a birth cohort (n = 109) Control: non-delinquents from a birth cohort (n = 545) Age: estimated mean = 13.5 years Matching: not matched Gender: 52% males Ethnicity: 100% Caucasian (New Zealand)	WISC R (VIQ)

Oosterlaan, et al. (2005)	N = 99 Disordered: BD youth from special schools (n = 61); ODD/CD (29.5%), ADHD (36.1%) & ODD/CD/ADHD (34.4%) Control: children from regular schools (n = 38) Age: mean = 10.3 years; range = 7 to 13 years Matching: age* Gender: 73.7% male Ethnicity: 100% Caucasian (Dutch)	semantic word fluency task & letter-word fluency task
Raine et al. (2005)	N = 325 Disordered: delinquents from a population-based sample (n = 169); ADHD (23%) Control: youth from a population-based sample (n = 156); ADHD (14.8%) Age: mean = 16.15; range = 16 to 17 years Matching: gender* Gender: 100% male Ethnicity: 58.8% African American & 41.2% Caucasian	WISC III (VIQ)
Speltz et al. (1999)	N = 160 Disordered: ODD youth from a psychiatric clinic (n = 80); ODD (28.8%), ODD/ADHD (56.3%), & ODD/other disorder	WPPSI-R (Comprehension), PPVT, & EOWPVT-R

	(15%) Control: recruited from the community (n = 80) Age: mean = 4.8 years; range = 3.9 to 5.7 years Matching: age, ethnicity, family structure, SES, & gender* Gender: 100% male Ethnicity: 81.25 % Caucasian & 18.75% African American	
Stellern et al. (1985)	N = 94 Disordered: BD youth at a residential school (n = 31) Control: public school students (n = 63) Age: mean = 10.5 years; range = 8 to 14 years Matching: not matched Gender: 62.8% male Ethnicity: 87.2% Caucasian	WISC R (VIQ)
Voorhees (1981)	N = 41 Disordered: delinquent youth in a correctional facility (n = 28) Control: high school students (n = 13) Age: mean = 15.5 years; range = 13 to 18 years Matching: NR	LNNB (Impressive Speech & Expressive Speech)

	Gender: 56% female	
	Ethnicity: NR	
Yeudall, et	N = 146	language modalities
al. (1982)	Disordered: delinquent youth at a residential facility (n = 99)	& oral word fluency
	Control: students in regular education (n = 47)	
	Age: mean = 14.7 years; range = 13 to 17 years	
	Matching: age*, sex*, handedness*	
	Gender: 63.7% male	
	Ethnicity: 100% Caucasian (Canadian)	

Note. Asterisks indicate that, although the groups were not matched during selection process, the groups did not differ significantly. ADHD = Attention-Deficit/Hyperactivity Disorder; ALQ = Adolescent Language Quotient; BD = Behavior Disorder; CD = Conduct Disorder; CELF = Clinical Evaluation of Language Function; E = Expressive; EOWPVT = Expressive One Word Picture Vocabulary Test; g = Hedges's *g* effect size statistic; G = Global; LNNB = Luria-Nebraska Neuropsychological Battery; n = number of participants per group; N = total sample size; NR = Not Reported; ODD = Oppositional Defiant Disorder; PPVT = Peabody Picture Vocabulary Test; R = Receptive; SD = standard deviation; SCOLP = Speed and Capacity of Language Processing; SE = standard error; SES = Socioeconomic Status; TLC = Test of Language Competence; TOAL = Test of Adolescent Language; WISC = Wechsler Intelligence Scale for Children.

Table 2.

Original Data Used in Calculating Effect Size

		Disordered Group			Control Group			
Author	Measure (Construct)	n	Mean	SD	n	Mean	SD	*g*
Cook et al. (1994)	WISC-R: Vocabulary (Global)	78	10.3	3.9	135	12.1	3.6	.48
Davis et al. (1991)	TOAL-2: ALQ (Global)	24	82.0	13.2	24	92.4	13.6	.76
	TOAL-2: RLQ (Receptive)	24	31.4	6.5	24	37.8	7.2	.92
	TOAL-2: ELQ (Expressive)	24	27.9	10.0	24	33.4	9.6	.55
Dery et al. (1999)	Aural Comprehension (Receptive)	59	16.5	1.0	29	16.8	1.1	.25
	Token Test (Receptive)	59	39.1	4.6	29	41.5	3.1	.57
	Visual Naming (Expressive)	59	40.2	5.3	29	43.0	5.1	.54
	Oral Word Association (Expressive)	59	22.3	7.0	29	25.9	5.6	.54
Giancola & Mezzich (2000)	TLC-E (Global)	223	88.1	14.3	97	100.2	14.7	.84

Golden & Golden (2001)	vocabulary task (Global)	15	26.0	6.5	15	40.7	8.0	1.97
	auditory comprehension task (Expressive)	15	11.3	1.8	15	.6	.7	7.3
Humber & Snow (2001)	SCOLP: Speed of Comprehension (Receptive)	15	30.6	17.3	15	54.3	10.4	1.62
	SCOLP: Spot the Word (Receptive)	15	39.1	4.2	15	42.3	3.9	.76
	TLC-E: Ambiguous Sentences (Receptive)	15	22.1	7.0	15	30.4	5.9	1.25
	TLC-E: Listening Comprehension (Receptive)	15	26.3	3.4	15	30.5	3.7	1.15
	TLC-E: Figurative Language (Receptive)	15	21.3	9.2	15	31.3	3.9	1.38
Karniski et al. (1982)	Auditory language function composite (Global)	54	6.6	1.3	51	7.	.8	1.01

Kusche et al. (1993)	WISC-R: Vocabulary (Global)	10 9	8.7	9.3	172	12.0	4.0	.50
	McCarthy Scales: Verbal Fluency (Expressive)	10 9	9.0	3.7	172	10.6	3.1	.48
Linz et al. (1990)	PPVT-R (Receptive)	20	82.3	12.7	20	95.1	13.4	.96
Miniutti (1991)	CELF-R: Total (Global)	27	62.3	13.2	26	83.3	14.9	1.47
	CELF-R: Receptive (Receptive)	27	67.0	12.1	26	87.6	13.5	1.59
	CELF-R: Expressive (Expressive)	27	63.3	14.1	26	81.5	15.9	1.20
Moffitt & Silva (1988b)	WISC-R: VIQ (Global)	10 9	98.1	16.1	545	105.0	14.1	.47
Oosterlaan et al. (2005)	Semantic word fluency (Expressive)	61	34.7	8.1	38	36.2	8.0	.19
	Letter-word fluency (Expressive)	61	14.8	4.7	38	16.8	6.6	.37
Raine et al. (2005)	WISC-III: VIQ (Global)	16 9	85.9	3.3	156	91.9	3.2	1.84
Speltz et al. (1999)	WISC-R: Comp. (Global)	80	11	3	80	12	3	.33

	PPVT (Receptive)	80	100	16	80	110	14	.66
	EOWPVT-R (Expressive)	80	99	18	80	117	17	1.02
Stellern et al. (1985)	WISC-R: VIQ (Global)	31	94.9	14.4	63	114.0	13.1	1.37
Voorhees (1981)	LNNB: Impressive (Receptive)	28	6.9	5.0	13	1.9	2.1	1.12
	LNNB: Expressive (Expressive)	28	6.5	5.6	13	1.5	1.5	1.03
Yeudall et al. (1982)	Halstead-Reitan: Language Modalities (Global)	99	5.0	3.0	46	2.2	1.5	1.06
	Halstead-Reitan: Oral Word Fluency (Expressive)	99	10.9	1.6	46	12.4	2.9	.71

Note. ALQ = Adolescent Language Quotient; CELF = Clinical Evaluation of Language Function; E = Expressive; ELQ = Expressive Language Quotient; EOWPVT = Expressive One Word Picture Vocabulary Test; g = Hedges's *g* effect size statistic; G = Global; LNNB = Luria-Nebraska Neuropsychological Battery; n = number of participants per group; PPVT = Peabody Picture Vocabulary Test; R = Receptive; RLQ = Receptive Language Quotient; SD = standard deviation; SE = standard error; SCOLP = Speed and Capacity of Language

Processing; TLC = Test of Language Competence; TOAL = Test of Adolescent Language; WISC = Wechsler Intelligence Scale for Children.

Table 3.

Descriptive Statistics of Participants for Each Meta-Analysis

			Age (years)		Gender (% male)		Ethnicity (% Caucasian)	
	N	Label (n)	range	mean	range	mean	range	mean
Global	4251	Del. (n = 7) CD (n = 4) ODD (n = 1) BD (n = 5)	4.8-16.6	12.8	0-100	68	13-100	77.7
Rec.	460	Del. (n = 7) CD (n = 4) ODD (n = 1) BD (n = 5)	4.8-16.6	13	44-100	76.7	13-100	31.5
Exp.	915	Del. (n = 7) CD (n = 4) ODD (n = 1) BD (n = 5)	4.8-16.6	11.6	44-100	73	13-100	79.1

Note. BD = Behavior Disorder; CD = Conduct Disorder; Del. = Delinquent; Exp. = Expressive; ODD = Oppositional Defiant Disorder; n = number of studies; N = number of participants; Rec. = Receptive.

Table 4.

Descriptive Statistics of Effect Size Distributions for Each Meta-Analysis

	n	N	Mean N	min	Q1	median	Q3	max	skew	SD	g
Global	17	4251	250	.27	.49	.89	1.23	1.97	.73	.12	.91
Rec.	8	460	65.7	.41	.79	.96	1.18	1.59	.04	.15	.92
Exp.	7	915	114.4	.28	.52	.63	1.02	1.20	.23	.11	.69

Note. Exp. = Expressive; g = Hedges's g effect size statistic; min = minimum; max = maximum; n = number of studies; N = number of participants, Q1 = 25th quartile; Q3 = 75th quartile; Rec. = Receptive; SD = standard deviation.

Table 5.

Global Language Meta-Analysis: Study-Level and Combined Results

Author	N	g	SE	Var.	IVW	RW	95% CI	p-value
Cook et al.	213	.48	.14	.02	4.47	6.79	.20, .77	.001
Davis et al.	48	.74	.29	.09	3.46	5.25	.17, 1.32	.012
Dery et al.	88	.47	.23	.05	3.93	5.96	.03, .92	.038
Giancola et al.	320	.84	.13	.02	4.57	6.95	.59, 1.08	.000
Golden & Golden	30	1.97	.44	.19	2.54	3.86	1.12, 2.83	.000
Humber & Snow	30	1.23	.39	.15	2.81	4.27	.47, 2.00	.002
Karniski et al.	105	1.01	.21	.04	4.08	6.19	.60, 1.41	.000
Kusche et al.	281	.49	.12	.02	4.58	6.96	.25, .73	.000
Linz et al.	40	.96	.33	.11	3.22	4.89	.32, 1.60	.003
Miniutti	53	1.42	.30	.09	3.39	5.14	.82, 2.01	.000
Moffitt & Silva	654	.474	.11	.01	4.67	7.10	.27, .68	.000
Oosterlaan et al.	99	.28	.21	.04	4.08	6.19	-.13, .68	.179
Raine et al.	325	1.84	.13	.02	4.54	6.90	1.6, 2.1	.000
Speltz et al.	160	.67	.16	.03	4.37	6.63	.35, .99	.000
Stellern et al.	94	1.37	.24	.06	3.85	5.84	.90, 1.84	.000
Voorhees	41	1.07	.35	.12	3.07	4.67	.39, 1.76	.002
Yeudall et al.	145	.89	.19	.03	4.22	6.41	.52, 1.25	.000
Combined	4251	.91	.12	.02	-	-	.67, 1.15	.000

Note. The inverse variance weight and relative weight only apply to study-level effect size values and do not apply to combined values; therefore, two cells in the Combined row were left empty. CI = confidence Interval; g = Hedges's g effect size statistic; IVW = inverse

variance weight; N = total sample size; RW = relative weight; SE = Standard Error; Var. = variance.

Table 6.

Receptive Language Meta-Analysis Study-Level and Combined Results

Author	N	g	SE	Var.	IVW	RW	95% CI	p-value
Davis et al.	48	.92	.30	.09	.592	13.70	.33,1.50	.002
Dery et al.	88	.41	.23	.05	7.61	17.60	-.04, .85	.075
Humber & Snow	30	1.23	.39	.15	4.30	9.96	.47, 2.0	.002
Linz et al.	40	.96	.33	.11	5.35	12.37	.32, 1.60	.003
Miniutti	53	1.59	.31	.10	5.68	13.13	.97, 2.20	.000
Speltz et al.	160	.66	.16	.03	9.46	21.88	.34, .98	.000
Voorhees	41	1.12	.35	.12	4.92	11.37	.43, 1.81	.001
Combined	460	.92	.15	.02	-	-	.62, 1.22	.000

Note. The inverse variance weight and relative weight only apply to study-level effect size values and do not apply to combined values; therefore, two cells in the Combined row were left empty. CI = confidence Interval; g = Hedges's g effect size statistic; IVW = inverse variance weight; N = total sample size; RW = relative weight; SE = Standard Error; Var. = variance.

Table 7.

Expressive Language Meta-Analysis Study-Level and Combined Results

Author	N	g	SE	Var.	IVW	RW	95% CI	p-value
Davis et al.	48	.55	.29	.08	7.56	9.16	-.02, 1.12	.057
Dery et al.	88	.54	.23	.05	9.92	12.02	.09, .98	.018
Kusche et al.	281	.48	.12	.04	15.58	18.88	.24, .72	.000
Miniutti	53	1.20	.30	.09	7.36	8.92	.62, 1.77	.000
Oosterlaan et al.	99	.28	.21	.04	10.96	13.28	-.13, .68	.179
Speltz et al.	160	1.02	.17	.03	13.04	15.80	.70, 1.35	.000
Voorhees	41	1.03	.35	.12	5.89	7.13	.34, 1.71	.000
Yeudall et al.	145	.71	.18	.03	12.21	14.79	.35, 1.07	.000
Combined	915	.69	.11	.01	-	-	.47, 91	.000

Note. The inverse variance weight and relative weight only apply to study-level effect size values and do not apply to combined values; therefore, two cells in the Combined row were left empty. CI = confidence interval; g = Hedges's g effect size statistic; IVW= inverse variance weight; N = total sample size; RW = relative weight; SE = standard error; Var. = variance.

Table 8.

Group Contrast Mixed Effects Moderator Analysis for Global Language
(groups determined by type of conduct problem)

Type of conduct Problem	n	*g*	SE	Var.	95% CI	p-value
BD	5	.88	.23	.05	.43, 1.32	.000
CD	4	.94	.22	.05	.50, 1.38	.000
Delinquent	7	1.03	.25	.06	.55, 1.52	.000
ODD	1	.28	.21	.04	-.13, .68	.18

Significance Test Results for Between Group Variance

	Q	p-value
	.93	.82

Note. BD = Behavior Disorder; CD = Conduct Disorder; CI = confidence interval; g = Hedges's g effect size statistic; n = number of studies; SE = standard error; Var. = variance CI = confidence Interval; g = Hedges's g effect size statistic; IVW = inverse variance weight; N = total sample size; RW = relative weight; SE = Standard Error; Var. = variance.

Figure 1.

Sampling Process for Article Procurement

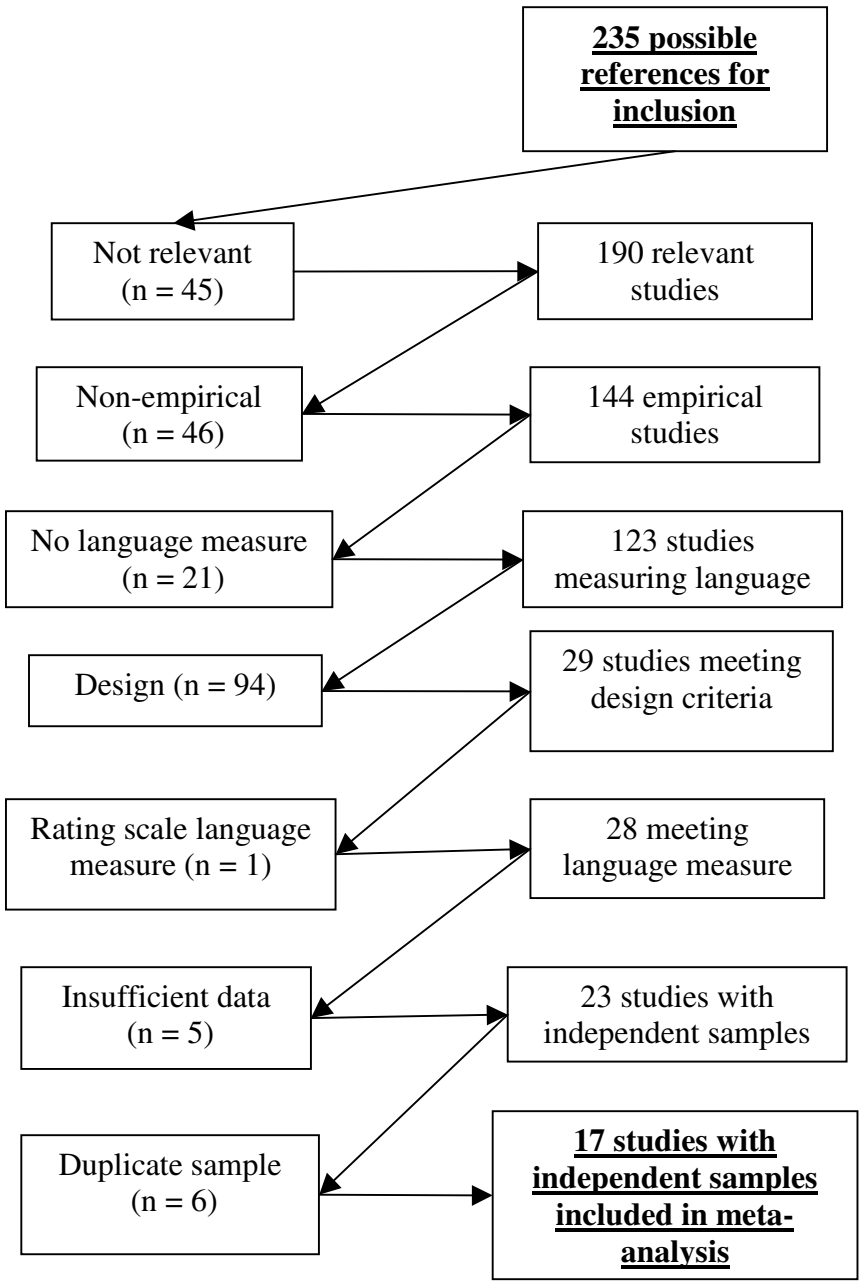


Figure 2.

Global Language Meta-Analysis: Study-Level and Combined Results

Study name	Hedges's g	Lower limit	Upper limit
Cook et al.	0.48	0.20	0.77
Davis et al.	0.74	0.17	1.32
Dery et al.	0.47	0.03	0.92
Giancola & Mezzich	0.84	0.59	1.08
Golden & Golden	1.97	1.12	2.83
Humber & Snow	1.23	0.47	2.00
Karniski et al.	1.01	0.60	1.41
Kusche et al.	0.49	0.25	0.73
Linz et al.	0.96	0.32	1.60
Miniutti	1.42	0.82	2.01
Moffitt & Silva	0.47	0.27	0.68
Oosterlaan et al.	0.28	-0.13	0.68
Raine et al.	1.84	1.59	2.10
Speltz et al.	0.67	0.35	0.99
Stellern et al.	1.37	0.90	1.84
Voorhees	1.07	0.39	1.76
Yeudall et al.	0.89	0.52	1.25
	0.91	0.67	1.15

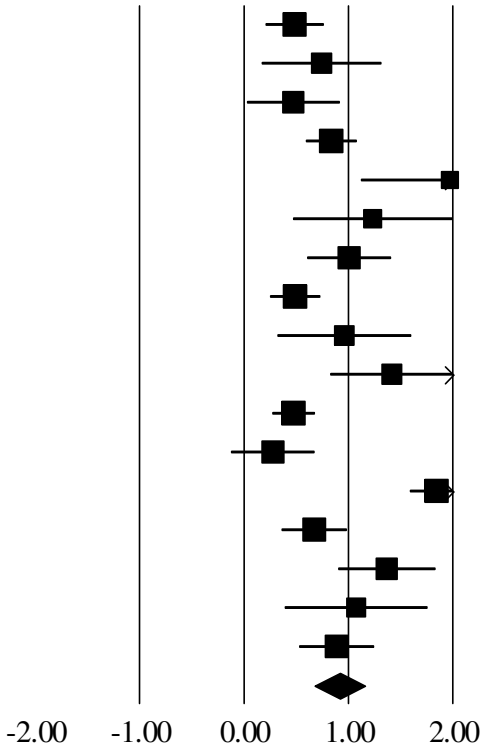


Figure 3.

Funnel Plot of Global Language Meta-Analysis:
Standard error of effect sizes displayed as a function of effect size (included studies only)

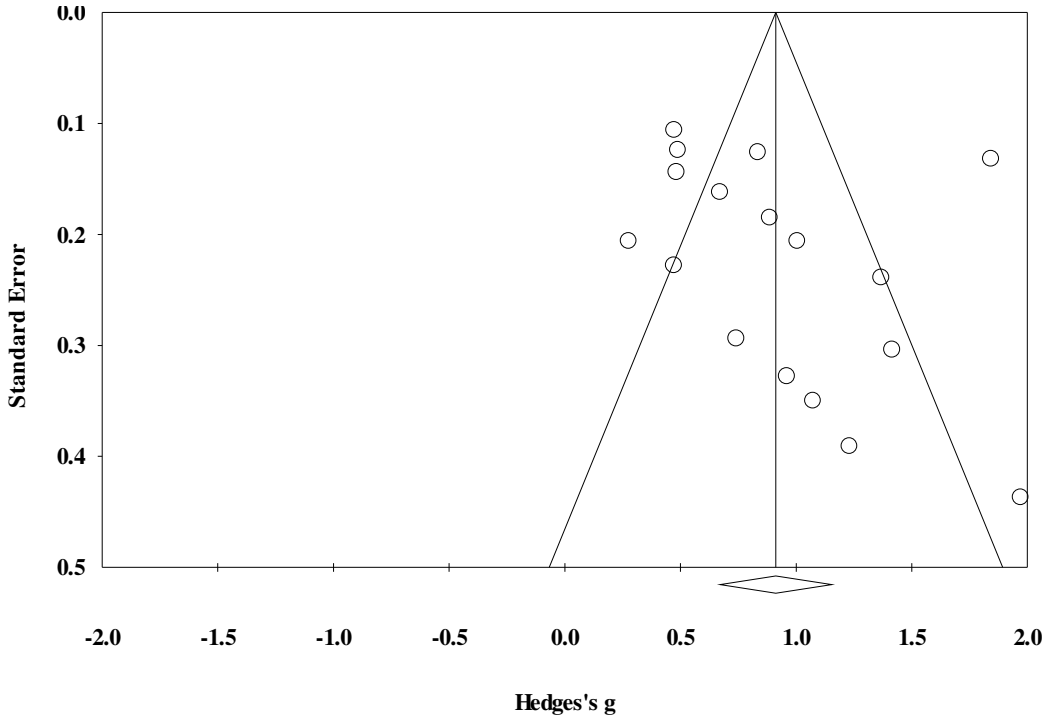


Figure 4.

Meta-Analysis of Receptive Language: Study-Level and Combined Results

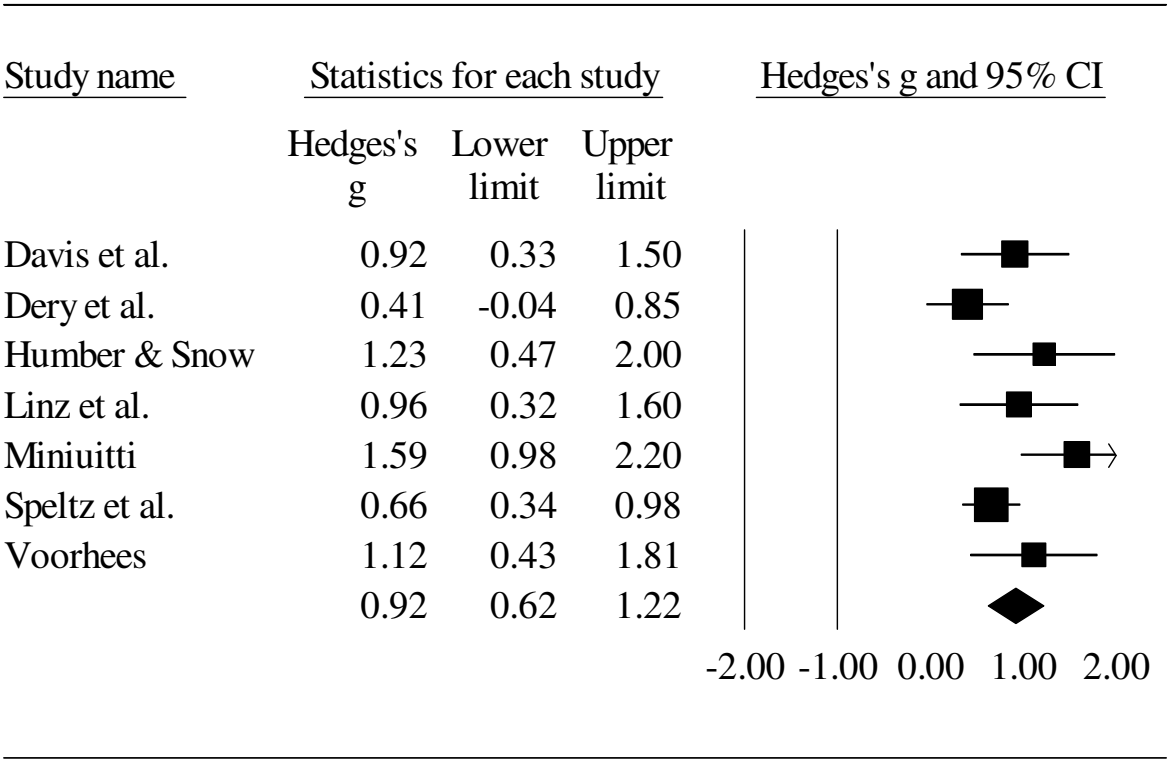


Figure 5.

Funnel Plot of Receptive Language Meta-Analysis:
Standard error of effect sizes displayed as a function of effect size (included studies only)

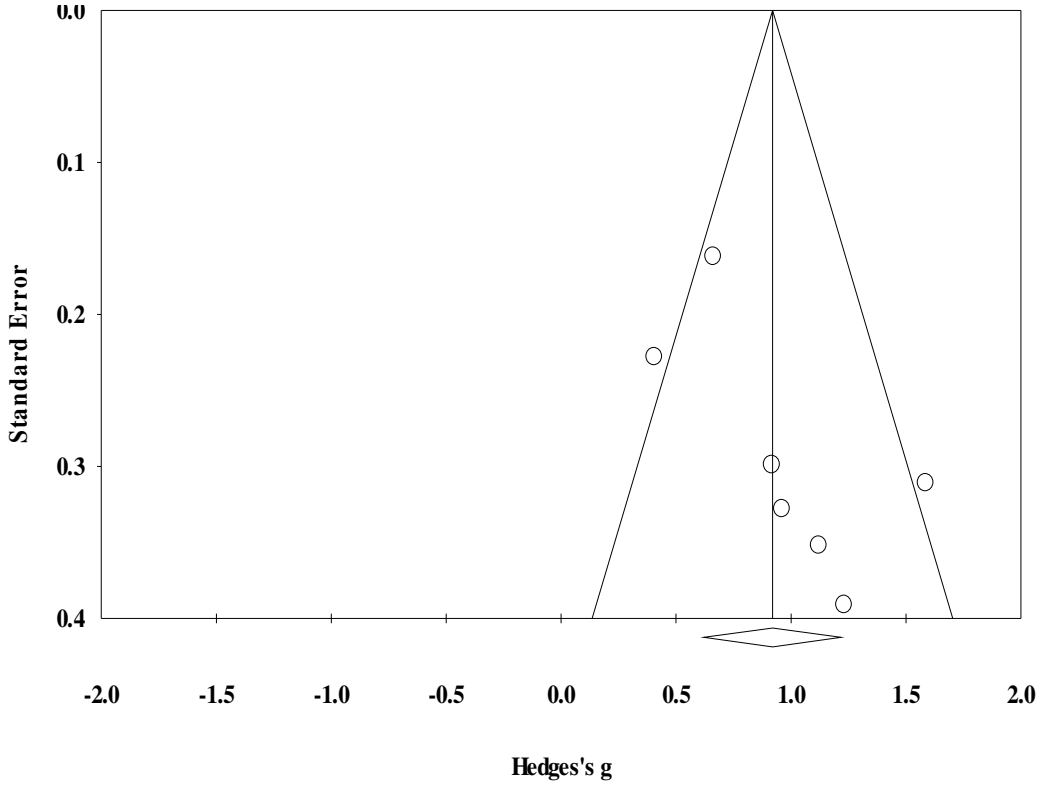


Figure 6.

Funnel Plot of Receptive Language Meta-Analysis:
Standard error of effect sizes displayed as a function of effect size (with imputed studies)

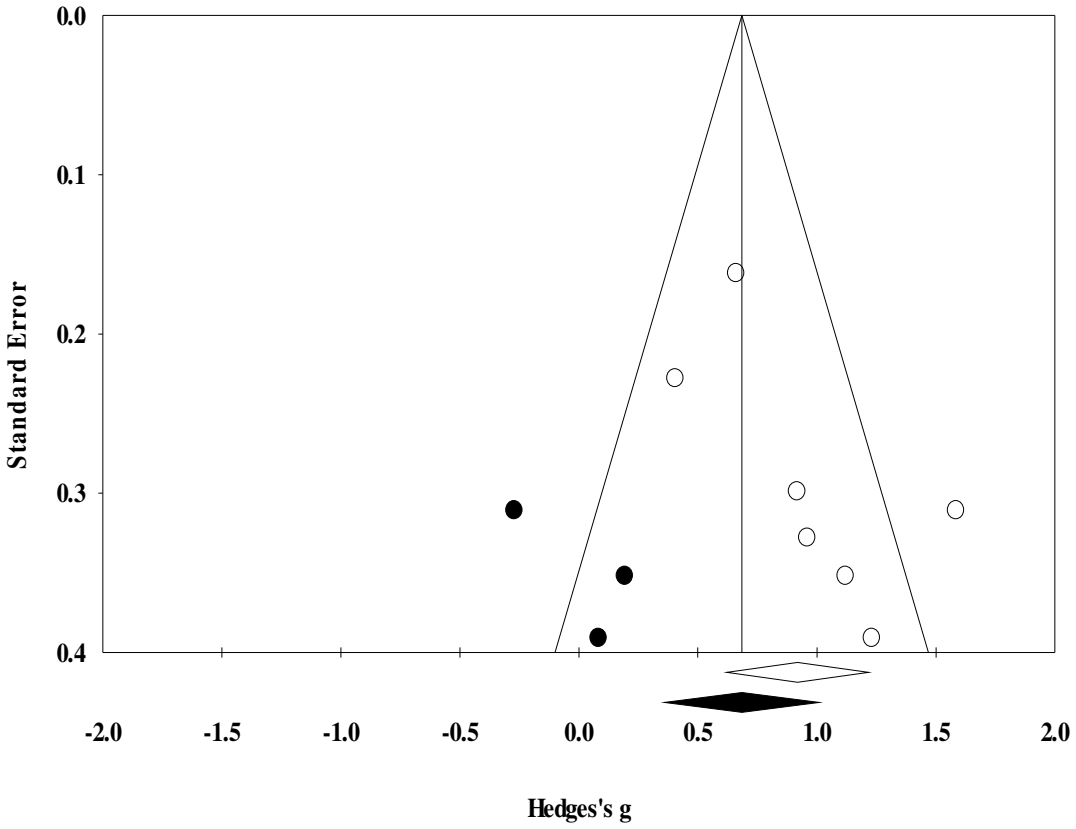


Figure 7.

Expressive Language Meta-Analysis: Study-Level and Combined Results

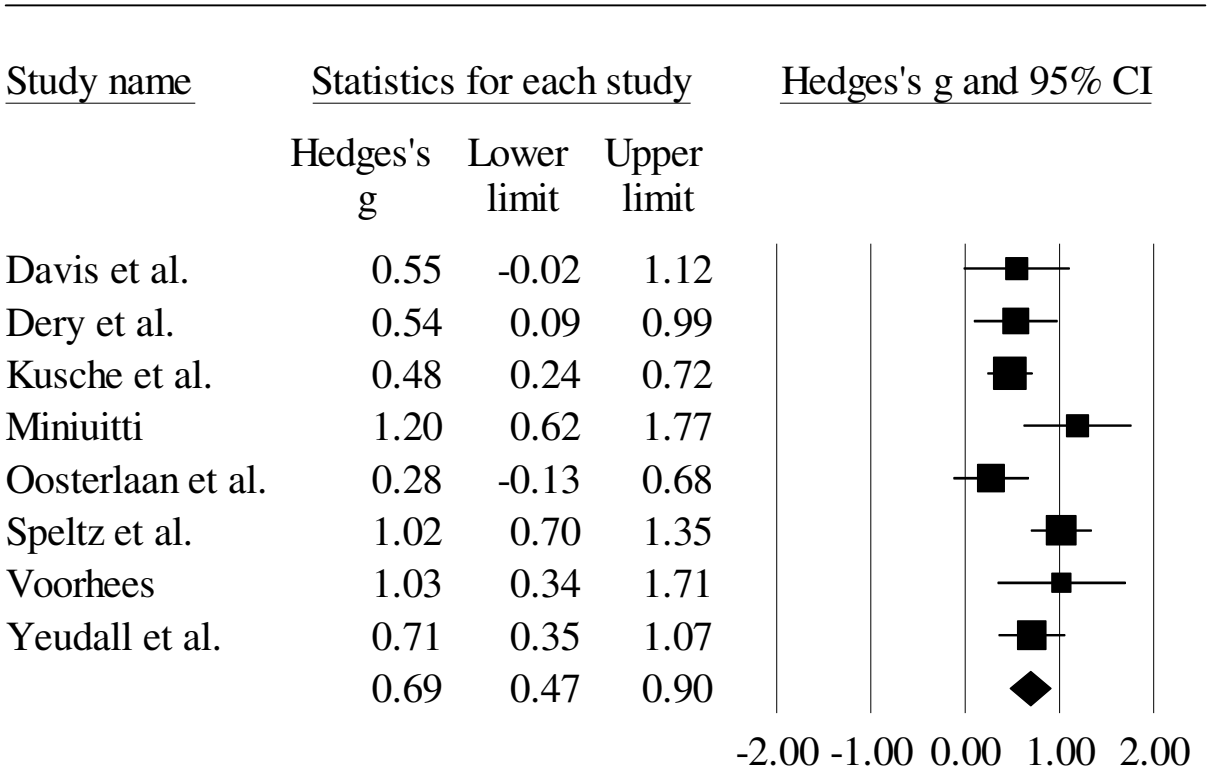


Figure 8.

Funnel Plot of Expressive Language Meta-Analysis:
Standard error of effect sizes displayed as a function of effect size (included studies only)

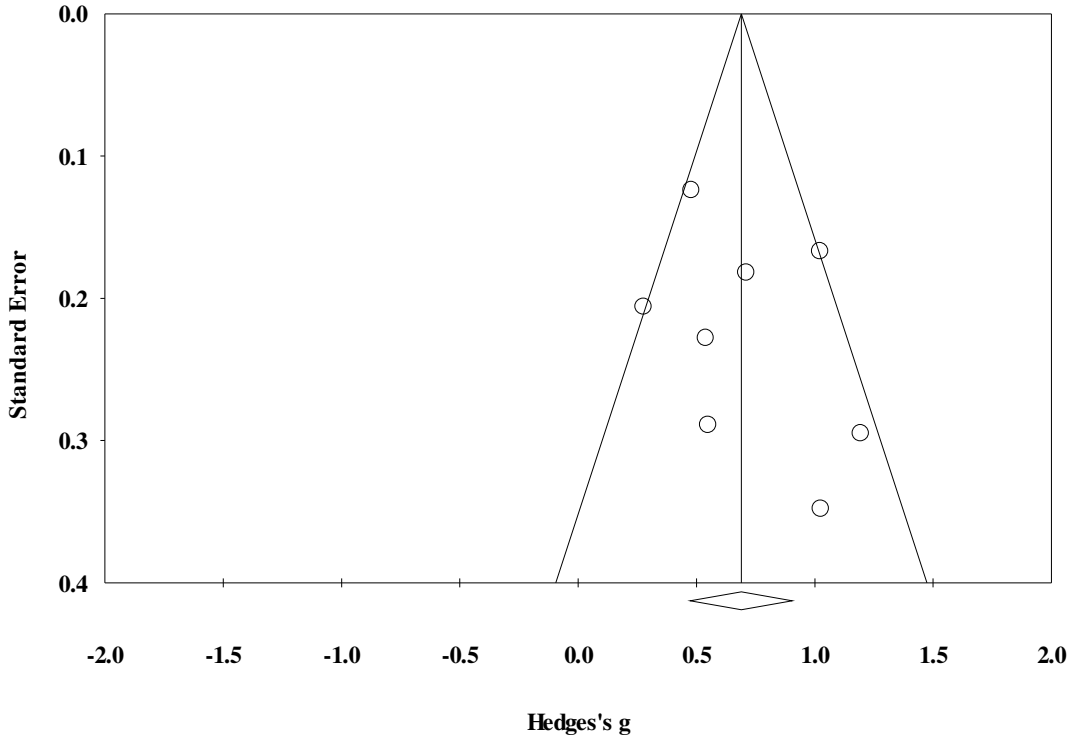


Figure 9.

Funnel Plot of Expressive Language Meta-Analysis:
Standard error of effect sizes displayed as a function of effect size (with imputed studies)

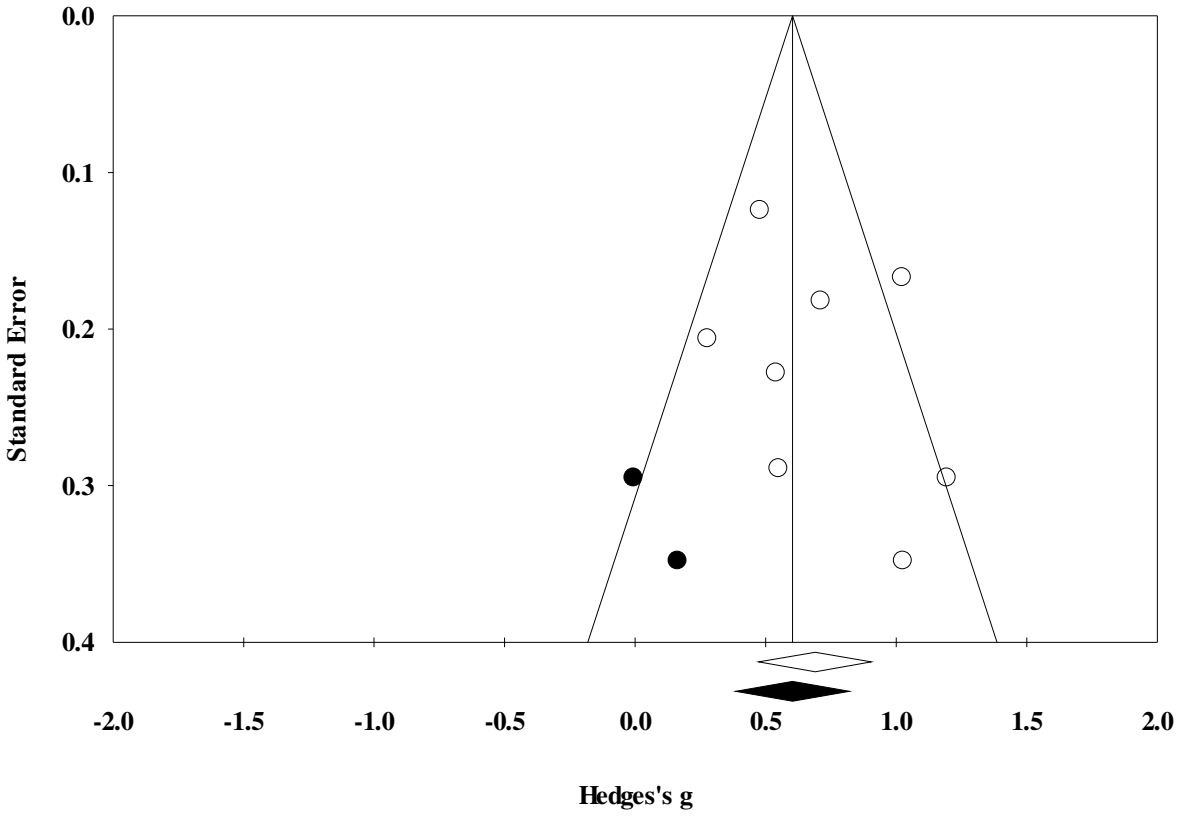


Figure 10.

Global Language Fixed Effects Meta-Regression:
Effect sizes displayed as a function of mean sample age (in years)

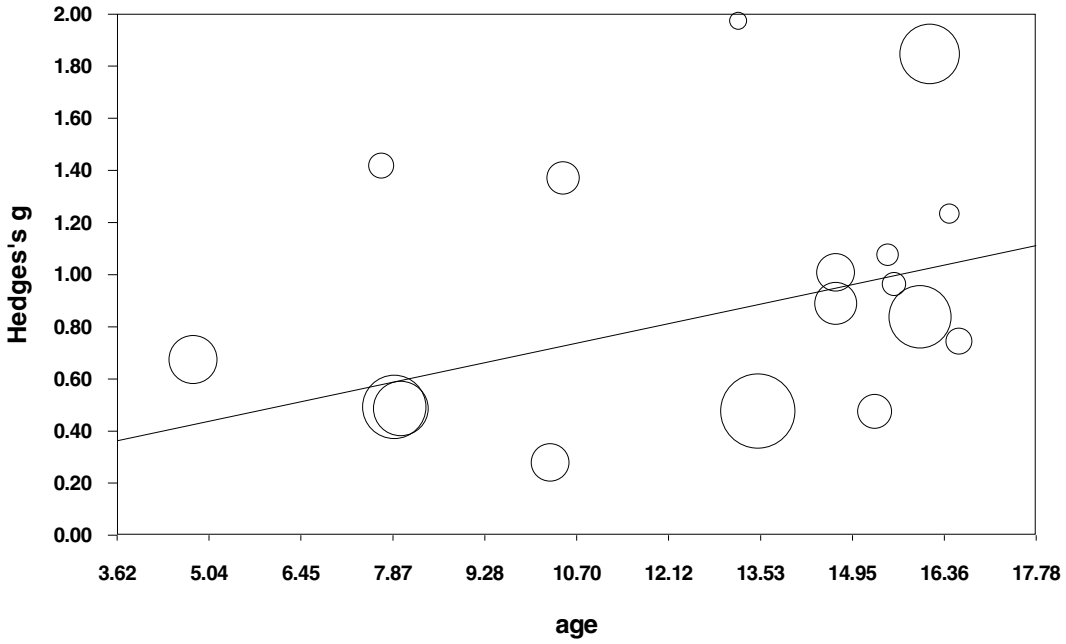


Figure 11.

Global Language Mixed Effects Meta-Regression:
Effect sizes displayed as a function of mean sample age (in years)

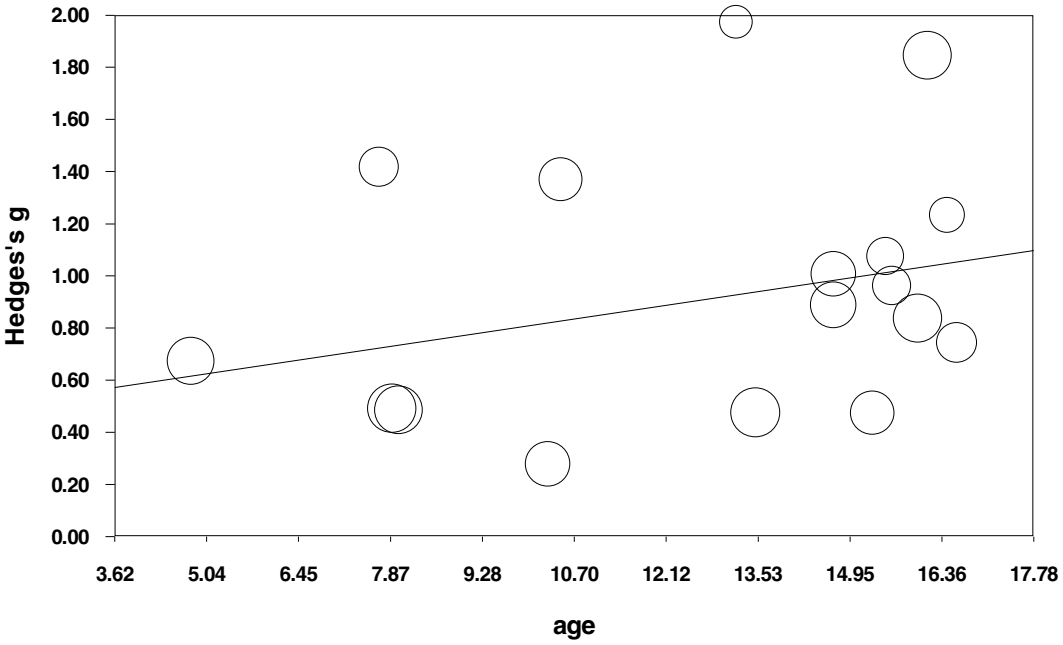


Figure 12.

Global Language Fixed Effects Meta-Regression:
Effect sizes displayed as a function of gender (percentage of males)

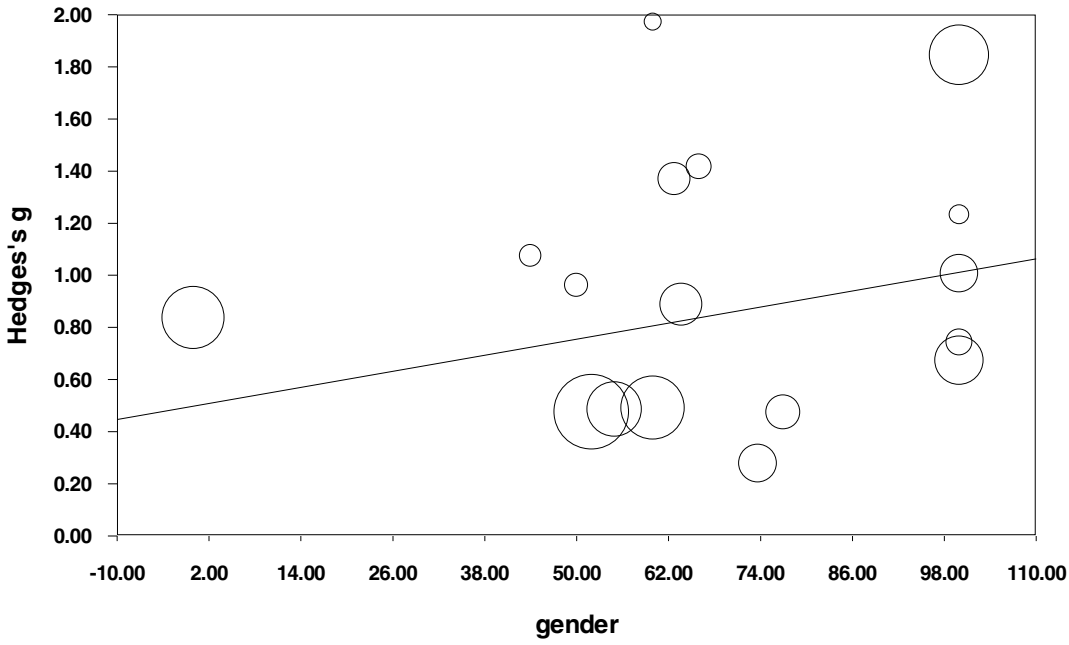


Figure 13.

Global Language Mixed Effects Meta-Regression:
Effect sizes displayed as a function of gender (percentage of males)

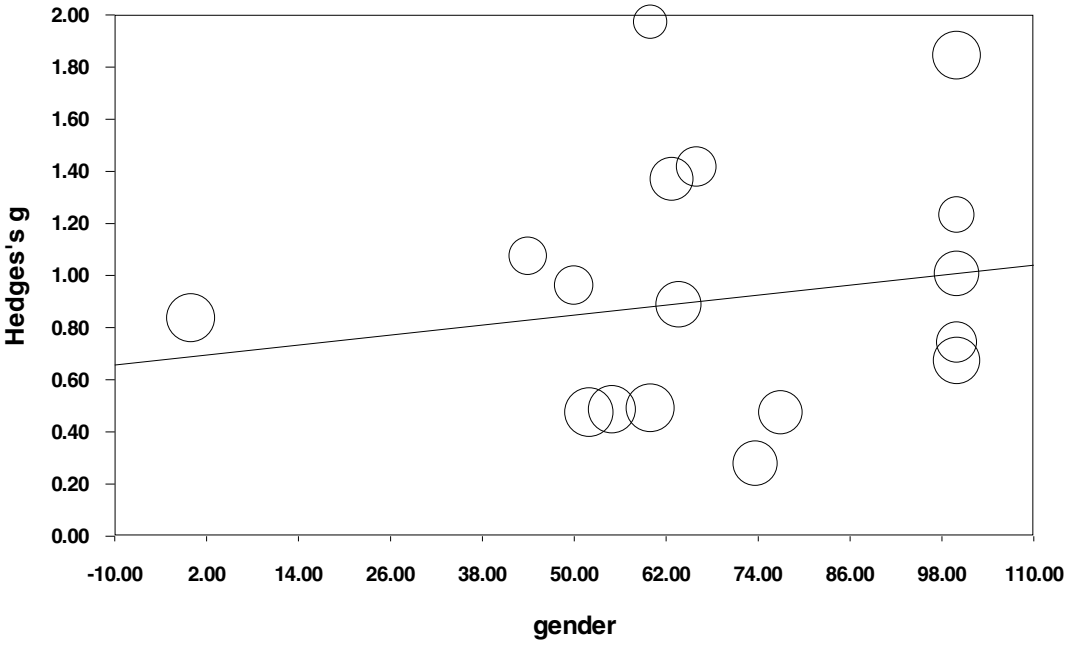


Figure 14.

Global Language Fixed Effects Meta-Regression:
Effect sizes displayed as a function of ethnicity (percentage of Caucasians)

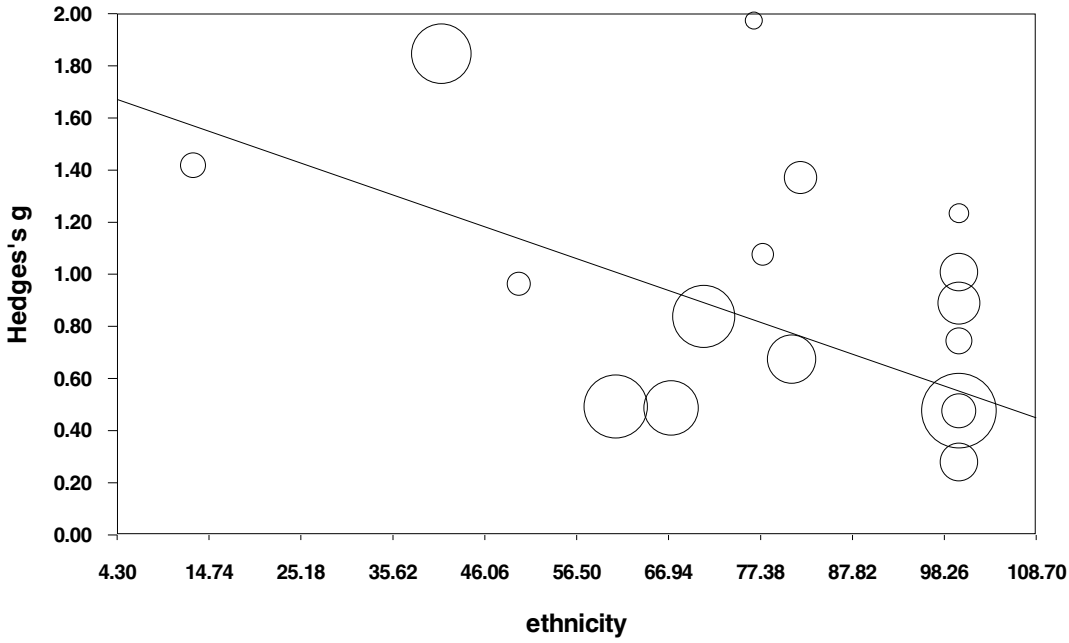


Figure 15.

Global Language Mixed Effects Meta-Regression:
Effect sizes displayed as a function of ethnicity (percentage of Caucasians)

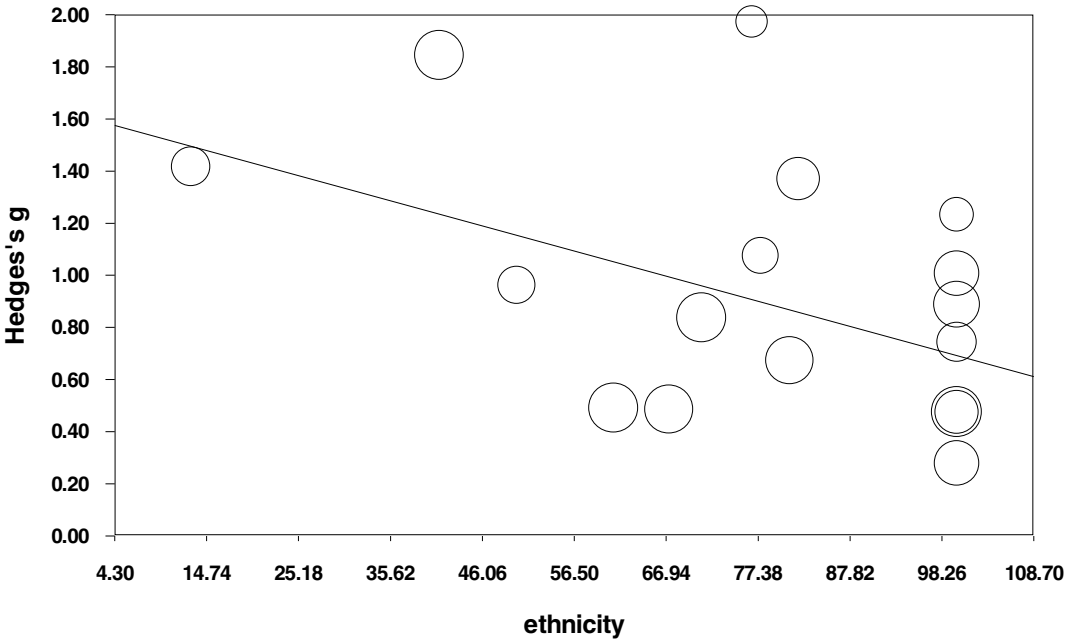


Figure 16.

Receptive Language Fixed Effects Meta-Regression:
Effect size displayed as a function of ethnicity (percentage of Caucasians)

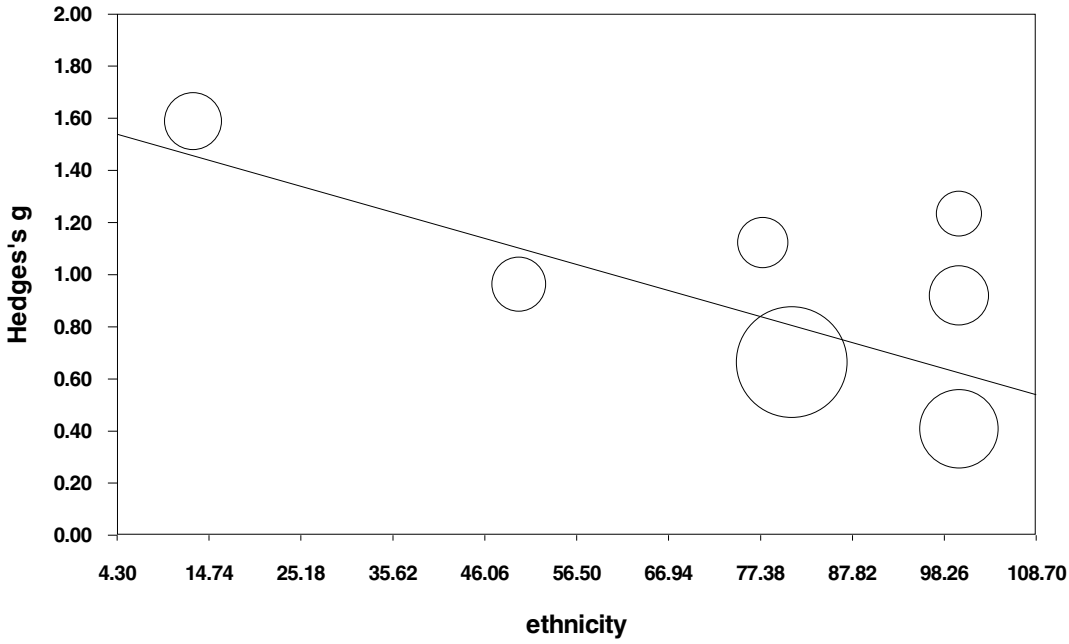


Figure 17.

Receptive Language Mixed Effects Meta-Regression:
Effect sizes displayed as a function of ethnicity (percentage of Caucasians)

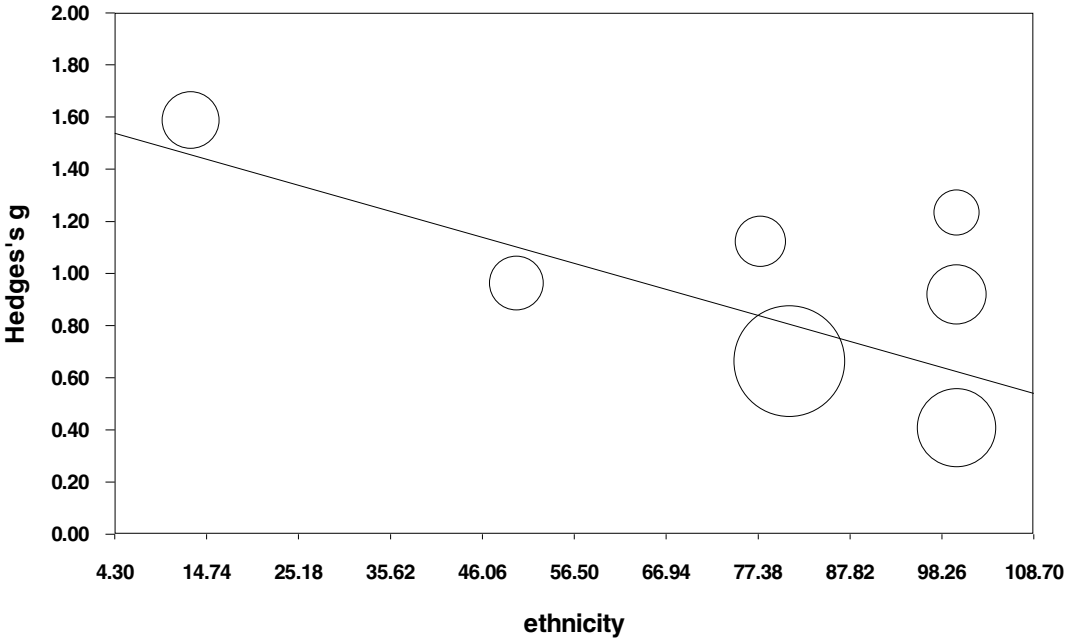


Figure 18.

Global Language Fixed Effects Meta-Regression:
Effect sizes displayed as a function of ADHD (percentage with ADHD

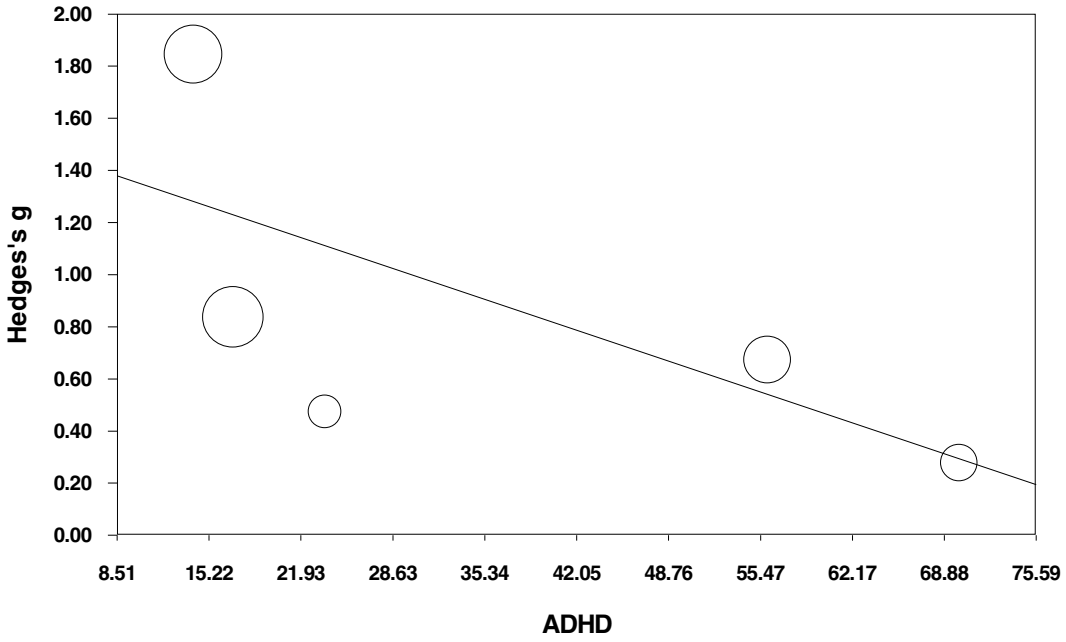
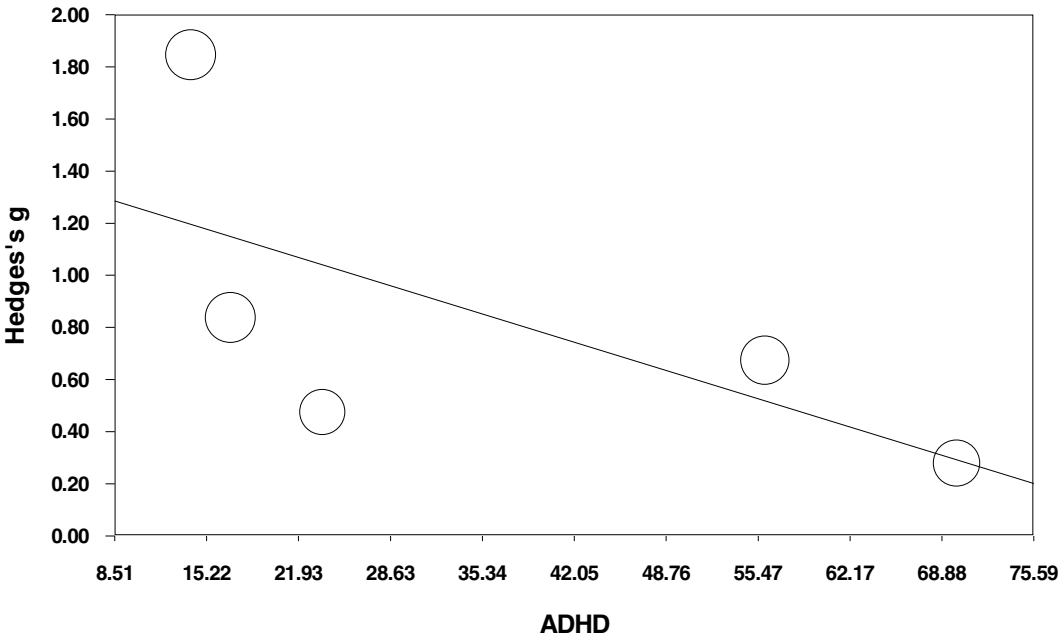


Figure 19.

Global Language Mixed Effects Meta-Regression:
Effect sizes displayed as a function of ADHD (percentage with ADHD)



APPENDIX A

DSM-IV-TR Diagnostic Criteria for Relevant Disorders (APA, 2000)

Conduct Disorder

A. A repetitive and persistent pattern of behavior in which the basic rights of others or major age-appropriate societal norms or rules are violated, as manifested by the presence of three (or more) of the following criteria in the past 12 months, with at least one criterion present in the past six months:

Aggression to people and animals

1. often bullies, threatens, or intimidates others
2. often initiates physical fights
3. has used a weapon that can cause serious physical harm to others
4. has been physically cruel to people
5. has been physically cruel to animals
6. has stolen while confronting a victim
7. has forced someone into sexual activity

Destruction of property

8. has deliberately engaged in fire setting with the intention of causing damage
9. has deliberately destroyed others' property (other than by fire setting)

Deceitfulness or theft

10. has broken into someone else's house, building, or car
11. often lies to obtain goods or favors or to avoid obligations
12. has stolen items or nontrivial value without confronting a victim

Serious violation of rules

13. often stays out at night despite parental prohibitions, beginning before age 13 years
14. has run away from home overnight at least twice while living in parental or parental surrogate home (or once without returning for a lengthy period)
15. is often truant from school, beginning before age 13 years

B. The disturbance in behavior causes clinically significant impairment in social, academic, or occupational functioning.

C. If the individual is age 18 year or older, criteria are not met for Antisocial Personality Disorder.

Note: CD is coded based on age of onset: Childhood Onset Type when at least one characteristic is present prior to 10 years of age, Adolescent Onset Type when there is an absence of criteria prior to 10 years of age, or Unspecified Onset when age is not known. CD is also coded according to severity: Mild, Moderate, or Severe.

<u>Oppositional Defiant Disorder</u>

A. A pattern of negativistic, hostile, and defiant behavior lasting at least 6 months, during which four (or more) of the following are present:

1. often loses temper
2. often argues with adults
3. often actively defies or refuses to comply with adults' requests or rules
4. often deliberately annoys people
5. often blames others for his or her mistakes or misbehavior
6. is often touchy or easily annoyed by others

7. is often angry and resentful
8. is often spiteful or vindictive

Note: In order for a criterion to be met, the behavior must occur more frequently than is typically observed in individuals of comparable age and developmental level.

B. The disturbance in behavior causes clinically significant impairment in social, academic, or occupational functioning.

C. The behaviors do not occur exclusively during the course of a Psychotic or Mood disorder.

D. Criteria are not met for CD, and, if the individual is age 18 years or older, criteria are not met for Antisocial Personality Disorder

Disruptive Behavior Disorder, Not Otherwise Specified

This category is for disorders characterized by conduct or oppositional defiant behaviors that do not meet the criteria for CD or ODD, but cause clinically significant impairment.

Attention-Deficit/Hyperactivity Disorder

A. Either (1) or (2):

1. six (or more) of the following symptoms of inattention have persisted for at least 6 months to a degree that is maladaptive and inconsistent with developmental level

Inattention

(a) often fails to give close attention to details or makes careless mistakes in schoolwork, work, or other activities

(b) often has difficulty sustaining attention in tasks or play activities

(c) often does not seem to listen when spoken to directly

(d) often does not follow through on instructions and fails to finish schoolwork, chores, or duties in the workplace (not due to oppositional behavior or failure to understand instructions)

(e) often has difficulty organizing tasks and activities

(f) often avoids, dislikes, or is reluctant to engage in tasks that require sustained mental effort (such as schoolwork or homework)

(g) often loses things necessary for tasks or activities (e.g., toys, school assignments, pencils, books, or tools)

(h) is often easily distracted by extraneous stimuli

(i) is often forgetful in daily activities

2. six (or more) of the following symptoms of hyperactivity-impulsivity have persisted for at least 6 months to a degree that is maladaptive and inconsistent with developmental level

Hyperactivity and Impulsivity

(a) often fidgets with hands or feet or squirms in seat

(b) often leaves seat in classroom or in other situations in which remaining seated is expected

(c) often runs about or climbs excessively in situations in which it is inappropriate (in adolescents or adults, may be limited to subjective feelings of restlessness)

(d) often has difficulty playing or engaging in leisure activities quietly

(e) is often "on the go" or often acts as if "driven by a motor"

(f) often blurts out answers before questions have been completed

(g) often had difficulty awaiting turn

(h) often interrupts or intrudes on others (e.g., butts into conversations or games)

B. Some hyperactive-impulsive or inattentive symptoms that caused impairment were present before the age of 7 years.

C. Some impairment from the symptoms is present in two or more settings (e.g., at school, work, and home)

D. There must be clear evidence of clinically significant impairment in social, academic, or occupational functioning.

E. The symptoms do not occur exclusively during the course of a Pervasive Developmental Disorder, Schizophrenia, or other Psychotic Disorder and are not better accounted for by another mental disorder (e.g., Mood Disorder, Anxiety Disorder, Dissociative Disorder, or a Personality Disorder).

Note: ADHD is coded based on type depending on symptom presentation: Combined Type if both Criterion A1 and A2 are met for the past 6 months, Predominantly Inattentive Type if Criterion A1 is met but Criterion A2 is not met for the past 6 months, and Predominantly Hyperactive-Impulsive Type if Criterion A2 is met but Criterion A1 is not met for the past 6 months.

<u>Expressive Language Disorder</u>

A. The scores obtained from standardized individually administered measures of expressive language development are substantially below those obtained from standardized measures of both nonverbal intellectual capacity and receptive language development. The disturbance may be manifest clinically by symptoms that include having a markedly limited vocabulary, making errors in tense, or having difficulty recalling words or producing sentences with developmentally appropriate length or complexity.

B. The difficulties with expressive language interfere with academic or occupational achievement or with social communication.

C. Criteria are not met for Mixed Receptive-Expressive Language Disorder or a Pervasive Developmental Disorder.

D. If Mental Retardation, a speech-motor or sensory deficit, or environmental deprivation is present, the language difficulties are in excess of those usually associated with these problems.

Mixed Receptive-Expressive Language Disorder

A. The scores obtained from standardized individually administered measures of both receptive and expressive language development are substantially below those obtained from standardized measures of both nonverbal intellectual capacity. Symptoms include those for Expressive Language Disorder as well as difficulty understanding words, sentences, or specific types of words, such as spatial terms.

B. The difficulties with receptive and expressive language significantly interfere with academic or occupational achievement or with social communication.

C. Criteria are not met for a Pervasive Developmental Disorder.

D. If Mental Retardation, a speech-motor or sensory deficit, or environmental deprivation is present, the language difficulties are in excess of those usually associated with these problems.

APPENDIX B

Reviewed Studies

Author	Design	Sample Description	Language Measure
Aronowitz et al. (1994)	Group Contrast	N = 20 (from inpatient psychiatric facility Age: mean = 15 years Group Comparisons: CD+ADHD (n = 9) vs. CD only (n =5); CD-positive (n = 14) vs. CD-negative (n = 6); ADHD-positive (n = 12) vs. ADHD-negative (n = 8) Gender: 12 males, 8 females	WISC R (VIQ)
Brickman et al. (1984)	Cor.	N = 71; Delinquents in a residential setting Age: mean = 16.3 years Ethnicity: 39 Caucasian, 32 minority	LNNB (Receptive Speech & Expressive Speech)
Camarata et al. (1988)	Cor.	N = 38; all in special education (SED) Age: mean = 10.9 years Gender: 30 male, 8 female	WPPSI (VIQ), WISC R (VIQ) TOLD-I (Listening & Speaking Composite)

Cole et al. (1993)	Group Contrast	N = 82 BD youth Groups: high vs moderate vs low risk Age: mean = 56.4 months Gender: 51 males, 31 females Ethnicity: 81.7% Caucasian	McCarthy Scales (Verbal Score) & FKSB (Verbal score)
Cornell & Wilson (1992)	Group Contrast	N = 149 delinquents Groups: Violent (n = 72) vs. nonviolent (n = 77) Age: mean = 15.2 years Gender: 145 males, 4 females Ethnicity: 105 minority, 44 Caucasian	WISC R & WAIS-R (VIQ)
Cook et al. (1994)	Group Contrast	N = 220 (full data for 213) Disordered group: students rated >1 SD above the mean on externalizing behavior Control: students rated < 1 SD above the mean on externalizing behavior Age: mean = 8.0 years Matching: not matched Gender: 55% male Ethnicity: 67.3% Caucasian, 24.5% African American, 8.2% other	WISC R (Vocabulary)

Coy et al. (2001)	Group Contrast	N = 168 Disordered: ODD (n = 88) Control: n = 80 Age: ranging from 3.8 to 5.5 years Gender: All male Ethnicity: 82% Caucasian, 18% minority	WPPSI-R (VIQ), PPVT, & EOWPVT
Culberton et al. (1989)	Factor Analysis	N = 82 delinquents from a correctional facility Age: ranging from 13 to 16 years Gender: 100% male	WISC R (VIQ)
Davis et al. (1991)	Group Contrast	N = 48 Disordered: institutionalized delinquents (n = 24) Control: high school students (n = 24) Age: mean = 16.6 years; Matching: age, FSIQ, gender*, & ethnicity* Gender: 100% male	TOAL-2
Dery et al. (1999)	Group Contrast	N = 88 Disordered: CD youth from various treatment settings (n = 59); ADHD (23.7%) Control: students in regular education (n = 29) Age: mean = 15.3; range = 13 to 17 years Matching: age*, gender*, SES*, & ethnicity* Gender: 100% male	Aural Comprehension of Words, Token Test, Visual Naming, & Controlled Oral Word

		Ethnicity: 100% Caucasian (French Canadian)	Association
Dishion et	Group	N = 70	Ammons Full-
al. (1984)	Contrast	Disordered: delinquent (n = 23)	Range Picture
		Control: nondelinquent (n = 47)	Vocabulary Test
		Age: 10th grade	
		Gender: 100% male	
		Ethnicity: 100% Caucasian	
Frost et al.	Group	N = 678 youth from the New Zealand	WISC-R (VIQ)
(1989)	Contrast	Dunedin birth cohort	
		Groups: Non-disordered (n = 605) vs. ADD	
		(n = 13) vs. CD (n = 17) vs. anxiety	
		(n = 14) vs. depression (n = 10) vs.	
		multiple disorders (n = 19)	
		Age: 13 years	
Giancola &	Group	N = 320	TLC-E (Total
Mezzich	Contrast	Disordered group: CD females from a variety	Score)
(2000)		of settings (n =22); ADHD (17%); anxiety	
		disorders (37%), Depression (35%),	
		Adjustment Disorder (10%), Dysthymia	
		(9%), eating disorder (8%), Bipolar	
		Disorder (.4%) & substance use (77%)	
		Control: obtained via a recruiting agency	
		(n = 97); Adjustment Disorder (4%)	

		Age: mean = 16.0 years	
		Matching: gender*	
		Gender: 100% female	
		Ethnicity: 71% Caucasian & 26% African American	
Grace & Sweeney (1986)	Group Contrast	N = 80 incarcerated delinquents African Americans: WISC-R (n = 20) vs. WAIS-R (n = 20) Caucasians: WISC-R (n = 20) vs. WAIS-R (n = 20) Age: mean = 16.0 years Ethnicity: 50% African American & 50% Caucasian	WISC R (VIQ) & WAIS-R (VIQ)
Golden & Golden (2001)	Group Contrast	N = 30 Disordered: CD youth from a mental health clinic (n = 15) Control: students in regular education (n = 15) Age: mean = 13.2 years Matching: not matched Gender: 60% males Ethnicity: 76.7% Caucasian, 10% African American, 10% Hispanic, & 3% other	vocabulary task & auditory comprehension task

Haynes & Bensch (1981)	Group Contrast	N = 90 adjudicated delinquents Groups: Recidivist (n = 54) vs. non-recidivist (n = 36) Age: ranging from 14 to 15 years Gender: 100% male Ethnicity: 100% Caucasian	WISC R (VIQ)
Haynes & Bensch (1983)	Group Contrast	N = 78 adjudicated delinquents Groups: Recidivist (n = 35) vs. non-recidivist (n= 43) Age: mean = 14.8 years Gender: 100% female Ethnicity: 100% Caucasian	WISC R (VIQ)
Hubble & Groff (1982)	Group Contrast	N = 55 incarcerated delinquents Age: NR Gender: 100% male Ethnicity: 100% Caucasian	WISC R (VIQ)
Humber & Snow (2001)	Group Contrast	N = 30 Disordered: adjudicated delinquents (n = 15) Control: public school students (n = 15) Age: mean = 16.45 years Matching: age*, gender* Gender: 100% male Ethnicity: 100% Caucasian (Australian)	SCOLP

Karniski et al. (1982)	Group Contrast	N = 105 Disordered: committed delinquents (n = 54) Control: middle/ high school students (n = 51) Age: mean = 14.7 years Matching: gender* & ethnicity* Gender: 100% male Ethnicity: 100% Caucasian	Auditory Language Function Composite
Kusche et al. (1993)	Group Contrast	N = 281 Disordered: BD students in special education (n = 109) Control: regular education students (n = 172) Age: mean = 7.9 years Matching: not matched Gender: 60% male Ethnicity: 61% Caucasian, 30% African American, 6% Asian, 3% Native American or Hispanic	WISC R (Vocabulary) & McCarthy Scales (Verbal Fluency)
Linz et al. (1990)	Group Contrast	N = 40 Disordered: CD youth in an evaluation center (n = 20) Control: from a variety of settings (n = 20) Age: mean = 15.6 years Matching: age, race, & gender*	PPVT-R

		Gender: 50% male Ethnicity: 50% Caucasian	
Mack & Warr-Leeper (1992)	Cor.	N = 20 psychiatric inpatients with severe BD Age: mean = 11.7 years Gender: 100% male Ethnicity: 19 Caucasian	WISC R (VIQ), CELF, TOAL, TLC, TOLD-I, EOWPVT, TOAL, PPVT-R, Token Test
McManas et al. (1985)	Cor.	N = 71 incarcerated delinquents Age: 16.2 years Gender: 40 males, 31 females Ethnicity: 39 Caucasian, 26 African American, 6 other	LNNB (Receptive Speech & Expressive Speech)
Miniutti (1991)	Group Contrast	N = 53 Disordered: BD students in special education (n = 27) Control: students in regular education (n = 26) Age: mean = 7.7 years Matching: age* Gender: 66% male Ethnicity: 76.9% African American, 13% Caucasian, & 10 % Hispanic	CELF-R (Total Composite, Receptive Composite, & Expressive Composite)

Moffitt (1988)	Group Contrast	N = 738 youth from the New Zealand Dunedin birth cohort Groups: delinquents vs non-delinquents Age: 13 years	WISC R (VIQ) & "Verbal Factor"
Moffitt (1990)	Group Contrast	N = 435 youth from the New Zealand Dunedin birth cohort Groups: no disorder (n = 348) vs. delinquent (n = 52) vs. ADD only (n = 16) vs. ADD/delinquent (n = 19) Age: 13 years Gender: 100% male	WISC R (VIQ)
Moffitt et al. (1981)	Group Contrast	N = 129 from a Danish birth cohort Groups: risk (n = 72) vs. control (n = 57): Age: NR	WISC R (VIQ)
Moffitt et al. (1994)	Cor.	N = unable to determine; from the New Zealand Dunedin birth cohort Age: 13 years Gender: 100% male	WISC R (VIQ) & Verbal Factor
Moffitt & Silva (1988a)	Group Contrast	N = 654 youth from the New Zealand Dunedin birth cohort Disordered group: delinquent youth (n = 109) Control: non-delinquent youth (n = 545)	WISC R (VIQ)

		Age: estimated mean = 13.5 years Matching: not matched Gender: 52% males Ethnicity: 100% Caucasian (New Zealand)	
Moffitt & Silva (1988b)	Group Contrast	N = 678 youth from the New Zealand Dunedin birth cohort Groups: detected delinquents (n = 40) vs. undetected delinquents (n = 40) vs. non-delinquents (n = 545) Age: 13 years	WISC R (VIQ)
Moffitt & Silva (1988c)	Group Contrast	N = 678 youth from the New Zealand Dunedin birth cohort Groups**:** delinquent (n = 124) vs. nondelinquent (n = 726) Age: 13 years	WISC R (VIQ) & Verbal Factor
Oosterlaan et al. (2005)	Group Contrast	N = 99 Disordered group: BD youth from special schools (n = 61); ODD/CD (29.5%), ADHD (36.1%) & ODD/CD/ADHD (34.4%) Control: youth from regular schools (n = 38) Age: mean = 10.3 years; range = 7 to 13 years Matching: age*	semantic word fluency task & letter-word fluency task

		Gender: 73.7% male Ethnicity: 100% Caucasian (Dutch)	
Petee & Walsh (1987)	Group Contrast	N = 125 delinquent youth on probation Groups: sample was split at the median VIQ-PIQ discrepancy; Small P > V discrepancy (n = 68) vs. large P > V discrepancy (n = 57) Gender: NR Ethnicity**:** 67 Caucasian, 58 African American	WISC R (VIQ)
Raine et al. (2005)	Group Contrast	N = 325 Disordered group: delinquents from a population-based sample (n = 169); ADHD (23%) Control: youth from a population-based sample (n = 156); ADHD (14.8%) Age: mean = 16.15; range = 16 to 17 years Matching: gender* Gender: 100% male Ethnicity: 58.8% African American & 41.2% Caucasian	WISC III (VIQ)
Robins et al. (1983)	Group Contrast	N = 50 adjudicated, non-incarcerated delinquents Groups: clinic-referred (n = 25) vs. non-clinic	WISC (VIQ) & WAIS (VIQ)

		(n = 25) Age: mean = 15.8 years Gender: 100% male Ethnicity: 42 Caucasian, 8 African American	
Sanger et al. (2001)	Correlational	N = 67 incarcerated delinquents Age: mean = 16.1 years Gender: 100% female Ethnicity: 41 Caucasian, 9 African American, 8 Hispanic, 7other	CELF-3 (Total, Receptive, & Expressive)
Schonfeld et al., (1988)	Group Contrast	N = 115 youth with CD; part of a birth cohort Groups: Sign positive (n = 58) vs. sign negative control group (n = 57) Age: 17 years Gender: 100% male Ethnicity: 100% African American	WAIS-R (VIQ)
Speltz et al. (1999)	Group Contrast	N = 160 Disordered: ODD youth from a psychiatric clinic (n = 80); ODD (28.8%), ODD/ADHD (56.3%), & ODD/other disorder (15%) Control: recruited from the community (n = 80) Age: mean = 4.8 years Matching: age, ethnicity, family structure,	WPPSI-R (Comprehension), PPVT, & EOWPVT-R

		SES, & gender* Gender: 100% male Ethnicity: 81.3 % Caucasian & 18.8% African American	
Stattin & Klackenberg -Larsson (1993)	Cor.	N = 122; from a birth cohort in Sweden Assessed beginning at age 3, through 17 Gender: 100% male	Language Development Composite & Westrin Intelligence Test
Stellern et al. (1985)	Group Contrast	N = 94 Disordered: BD youth at a residential school (n = 31) Control: public school students (n = 63) Age: mean = 10.5 years; range = 8 to 14 years Matching: not matched Gender: 62.8% male	WISC R (VIQ)
Tarter et al. (1983)	Cor.	N = 73 delinquents referred for neuropsychological evaluation by juvenile court Age: mean = 15.6 years Gender: 100% male Ethnicity: 43 Caucasian, 30 African American	WISC R/WAIS-R (VIQ), Token Test, & PPVT
Tarter et al.	Group	N = 101 delinquents	WISC-R/WAIS-

(1984)	Contrast	Groups: abused (n = 27) vs. non-abused (n = 74) Age: mean = 15.7 years Gender: 82% male Ethnicity: 66 Caucasian, 53 African American	R (VIQ), Token Test, & PPVT
Tarter et al. (1985)	Group Contrast	N = 101 juvenile delinquents Groups: VIQ > PIQ (n = 8) vs. PIQ > VIQ (n = 29) vs. VIQ = PIQ (n = 64) Age: 15 years	WISC R/WAIS-R (VIQ)
Teichner et al. (2000)	Cluster Analysis	N = 77 youth with CD (75%) & ODD (17%) Age: mean = 15.3 years	LNNB-III
Tramontana & Hooper (1987)	Group Contrast	N = 50 psychiatric inpatients Groups: CD (n = 17) vs. Depression (n = 17) vs. TB I (n = 15) Age: mean = 14.99 years Gender: 36 males, 14 females	LNNB
Voorhees (1981)	Group Contrast	N = 41 Disordered: delinquents in a correctional facility (n = 28) Control: high school students (n = 13) Age: mean = 15.5 years Matching: NR Gender: 56% female	LNNB

		Ethnicity: NR	
Walsh & Beyer (1986)	Group Contrast	N = 131 juvenile delinquents on probation Groups: small PIQ > VIQ discrepancy (n = 89) vs. large PIQ > VIQ (n = 42) Age: NR	WISC R (VIQ)
Warr-Leeper et al. (1994)	Cor.	N = 20 in residential treatment placements Age: mean = 11.8 years Gender: 100% male Ethnicity: 100% Caucasian (Canadian)	WISC R (VIQ), TOAL, TLC
Werry et al. (1987)	Group Contrast	N = 95 Groups: ADHD (n = 39) vs. ADHD and CD/ODD (n = 35) vs. Anxiety (n = 21) Age: ranging from 5 to 13 years	PPVT
Wolff et al. (1982)	Group Contrast	N = 152 Disordered: delinquents (n = 56) Control: lower-middle class control (n = 48) & upper-middle class control (n = 48) Age: ranging from 14 to 16 years Gender: 100% male Ethnicity: 100% Caucasian	PPVT, Token Test, & Boston Naming Test
Wong & Cornell (1999)	Cor.	N = 95 adjudicated delinquents Age: mean = 16.2 years Gender: 100% male	WISC R/WISC-III/WAIS-R (VIQ)

		Ethnicity: 71.6% minority	
Yeudall et al. (1982)	Group Contrast	N = 146 Disordered: delinquents at a residential facility (n = 99) Control: students in regular education (n = 47) Age: mean = 14.7 years Matching: age*, sex*, handedness* Gender: 63.7% male Ethnicity: 100% Caucasian (Canadian)	language modalities & oral word fluency
Zincus & Gottlieb (1983)	Cor.	N = 30 institutionalized delinquents Age: ranging from 13 to 18 years Gender: 100% male Ethnicity: 60% African American	WISC R/WAIS-R (VIQ) & PPVT

Note. Asterisks indicate that, although the groups were not matched during selection process, the groups did not differ significantly. ADHD = Attention-Deficit/Hyperactivity Disorder; ALQ = Adolescent Language Quotient; BD = Behavior Disorder; CD = Conduct Disorder; CELF = Clinical Evaluation of Language Function; Cor. = Correlational; E = Expressive; EOWPVT = Expressive One Word Picture Vocabulary Test; g = Hedges's *g* effect size statistic; G = Global; LNNB = Luria-Nebraska Neuropsychological Battery; n = number of participants per group; N = total sample size; NR = Not Reported; ODD = Oppositional Defiant Disorder; PPVT = Peabody Picture Vocabulary Test; R = Receptive; SD = standard deviation; SCOLP = Speed and Capacity of Language Processing; SE = standard error; SES

= Socioeconomic Status; TLC = Test of Language Competence; TOAL = Test of Adolescent Language; WISC = Wechsler Intelligence Scale for Children.

APPENDIX C

Coding Manual

Report Identification

1. Unique ID: Record the ID number printed in the top right corner of each study.
2. Year: Record the 2-digit year of publication.

Study Characteristics

3. Sample source (from what type of setting):
    - Disordered group: specify a) inpatient/residential vs. outpatient b) psychiatric/clinic vs. correctional/criminal records vs. school vs. combination c) part of a birth cohort d) other relevant information or if cannot be determined
    - Control group: specify a) community vs. school vs. psychiatric/clinic vs. combination b) part of a birth cohort c) other relevant information or cannot be determined

    Note: this item will be post-coded because studies use a large range of descriptors, which make it difficult to determine the most appropriate coding scheme a priori.
4. Mean sample age: Record the mean age for the entire sample at the time of assessment, to one decimal place. If the study only reports grade, record age according to the following guidelines: Kindergarten = 5.0 First grade = 6.0, Second grade = 7.0, Third grade = 8.0, etc. Consider that some studies report age in terms of number of years and months (i.e. 10-2 indicates 10 years 2 months) whereas other studies report age as an integer (i.e. 10.2 indicates 10 and 2/10 years. Regardless of how the study reports age, record age in the form of an integer.

5. Standard deviation of age: Report for the entire sample at the time of assessment, to one decimal place.
6. Gender: Record % of males in total sample.
    1. <33% male
    2. From 33.1% up to, but not including, 66%
    3. >66% male
7. Ethnicity: Record % of participants representing each ethnic group.
    1. >60% White
    2. >60% African American
    3. >60% Hispanic
    4. >60% other minority (specify)
    5. Mixed, cannot estimate proportion
8. SES: Transcribe information that pertains to SES (type/name of index, exact value/rating if applicable, income cut-off ranges, or other relevant information). Note: this item will be post-coded because studies use a range of methods to determine SES, which makes it difficult to determine the most appropriate coding scheme a priori.
9. Label: Record the label used to describe the disordered group.
    1. Delinquent
    2. CD
    3. ODD

4. BD: this includes SED (or any state-specific equivalent label), disruptive behavior disorder, behaviorally disturbed, or any other similar general description
5. Combination of two or more of the above labels (specify which ones)

Methodology

10. Matching: On what variables were the groups matched? (Record all applicable numbers)
    1. Age
    2. Gender
    3. Ethnicity
    4. SES
    5. Other (specify which variables)
    6. Groups not matched
11. Did researchers assess for comorbid ADHD? Y / N
12. Percentage of the disordered group with comorbid ADHD
13. What method did researchers use to reach a diagnosis of ADHD?
    1. Structured interview based on DSM (III or IV) criteria (specify)
    2. Semi-structured interview based on DSM (III or IV) criteria (specify)
    3. Rating scale or checklist (specify)
    4. Multiple methods (specify)
    5. Other (specify)
    6. Comorbidity of ADHD/ADD not determined/reported

Effect Size Information

14. Was an ES reported in the study? Y / N (If No, skip to # 16)
15. If yes to #13, record the value of the ES. Indicate the direction of effect by noting which group performed better on the outcome measure. On some measures, low scores, versus high scores, indicate better performance. Record the page number on which the ES is found.
16. If yes to #13, record the type of summary statistics from which the ES was derived. Indicate the page number in the study where this information can be found.
17. If no to #13, transcribe information from which an ES can be calculated. Indicate the page number where this information can be found. Use the following hierarchy to determine the best information for transcription with 1 being the most preferred to 5 being the least preferred:
    1. Mean, SD, and sample size for each group
    2. Descriptive data from which mean and SD can be computed
    3. Significance tests (t-values, df, and sample sizes; F-values, df, and sample sizes)
    4. Significance levels (an exact p value for a t-test or one-way ANOVA along with sample sizes)
    5. Indicate (by writing in the same space provided for transcription of ES information) if the study does not report enough information from which to calculate an ES
18. Record the actual number of subjects providing ES information, if cases are lost.

19. Record the name of the measure used to determine language functioning. Write the full name of the measure, including the edition. Also indicate if a short form was used. If the study used a composite, indicate what tasks/subtests/tests were used in reaching the composite.
20. Is reliability reported for this measure? Y / N
21. Record the estimated reliability
22. Record the type of reliability
    1. Alpha
    2. Internal consistency
    3. Kappa
    4. Percent agreement
    5. Split half
    6. Test-retest
    7. Other (specify)

APPENDIX D

Coding Protocol: Study-Level

_____	1. Unique id
_____	2. Year
	3a. Sample Source: Disordered Group (transcription and page)
	3b. Sample Source: Control Group (transcription and page)
_____	4. Mean sample age
_____	5. SD of sample age
_____	6. Gender
_____	7. Ethnicity
	8. SES (transcription and page)
_____	9. Label used to describe disordered group
_____	10. Matching variables
Y / N	11. Was comorbid ADHD assessed?
_____	12. Percentage of disordered group with ADHD
_____	13. Method used to reach diagnoses

APPENDIX E

Coding Protocol: Effect Size-Level

_____	Unique ID
Y / N	1. ES reported? (if no, skip to #16)
_____	2. Value of ES and page
	3. Summary statistics and page
	4. Transcribe info from which ES can be calculated and page
_____	5. Number of subjects providing ES info
	6. Name of measure
Y /N	7. Reliability reported?
_____	8. Reliability
_____	9. Type of reliability

REFERENCES

References marked with an asterisk indicate studies included in the meta-analysis.

Alvarez, H.K. & Ollendick, T. H. (2003). Individual and psychosocial risk factors. In C.A. Essau (Ed.), *Conduct and oppositional defiant disorders: Epidemiology, risk factors, and treatment* (pp. 97-116). New Jersey: Lawrence Erlbaum Associates.

American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision*. Washington, D.C., American Psychiatric Association, 2000.

Angold, A. & Costello, E. (2001). The epidemiology of disorders of conduct: Nosological issues and comorbidity. In J. Hill & B. Maughan (Eds.), *Conduct disorders in childhood and adolescence* (pp. 126-168). New York, New York: Cambridge University Press.

Aronowitz, B., Leibowitz, M., Hollander, E., Fazzini, E., Durlach-Misteli, C., Frenkel, M., et al. (1994). Neuropsychiatric and neuropsychological findings in conduct disorder and attention-deficit hyperactivity disorder. *Journal of Neuropsychiatry, 6*, 245-249.

Ayllon, T. & Roberts, M. (1974). Eliminating discipline problems by strengthening academic performance. *Journal of Applied Behavior Analysis*, 71-76.

Baker, L. & Cantwell, D.P. (1982). Developmental, social, and behavioral characteristics of speech and language disordered children. *Child Psychiatry and Human Development, 12* (4), 195-206.

Baker, L., Cantwell, D.P., & Mattison, R.E. (1980). Behavior problems in children with pure speech disorders and in children with combined speech and language disorders. *Journal of Abnormal Child Psychology, 8*(2), 245-256.

Beitchman, J.H., Hood, J., Rochon, J., & Peterson, M. (1989). Empirical classification of speech/language impairment in children II: Behavioral characteristics. *Journal of the American Academy of Child and Adolescent Psychiatry, 28*(1), 118-123.

Beitchman, J.H., Nair, R., Clegg, M., Ferguson, B., & Patel, P.G. (1986). Prevalence of psychiatric disorders in children with speech and language disorders. *Journal of the American Academy of Child Psychiatry, 25*(4), 528-535.

Beitchman, J.H., Wilson, B., Johnson, C.J., Atkinson, L., Young, A., Adlaf, E., et al.

(2001). Fourteen-year follow-up of speech/language-impaired and control children: Psychiatric outcome. *Journal of the American Academy of Child and Adolescent Psychiatry, 40*(1), 75-82.

Benasich, A.A., Curtiss, S., & Tallal, P. (1993). Language, learning, and behavioral disturbances in childhood: A longitudinal perspective. *Journal of the American Academy of Child and Adolescent Psychiatry, 32*(3), 585-595.

Berman, A. & Siegal, A.W. (1976). Adaptive and learning skills in juvenile delinquents: A neuropsychological analysis. *Journal of Learning Disabilities, 9*(9), 583-590.

Borenstein, M., Hedges, L., Higgins, J. & Rothstein, H. (2000). Comprehensive meta-analysis (Version 2) [Computer software and manual]. Englewood, NJ: Biostat.

Brickman, A.S., McManus, M., Grapentine, W.L., & Alessi, N. (1984). Neuropsychological assessment of seriously delinquent adolescents. *Journal of the American Academy of Child Psychiatry, 23*(4), 453-457.

Buikhuisen, W., Bontekoe, E.H.M., Plas-Korenhoff, C.v.d., & Meijs, B.W.G.P. (1988). Biological, psychological, and social factors related to juvenile delinquency. In W. Buikhuisen and S.A. Mednick (Eds.), *Explaining Criminal Behavior: Interdisciplinary Approaches* (pp. 121-140). New York: E.J. Brill.

Camarata, S.M., Hughes, C.A., & Ruhl, K.L. (1988). Mild/moderate behaviorally disordered students: A population at risk for language disorders. *Language, Speech and Hearing Services in Schools, 19,* 191-200.

Camp, B.W., Zimet, S.G., van Doorninck, W.J., & Dahlem, N.W. (1977). Verbal abilities in young aggressive boys. *Journal of Educational Psychology, 69*(2), 129-135.

Cantwell, C.P. & Baker, L. (1980). Psychiatric and behavioral characteristics of children with communication disorders. *Journal of Pediatric Psychology, 5*(2), 161-178.

Cantwell, C.P. & Baker, L. (1985). Psychiatric and learning disorders in children with speech and language disorders: A descriptive analysis. *Advances in Learning and Behavioral Disabilities, 4*, 29-47.

Carson, D.K., Klee, T., Perry, C.K., Donaghy, T., & Muskina, G. (1997). Measures of language proficiency as predictors of behavioral difficulties, social, and cognitive development in 2-year-old children. *Perceptual and Motor Skills, 84*, 923-930.

Carson, D.K., Klee, T., Perry, C.K., Muskina, G., & Donaghy, T. (1998). Comparisons of children with delayed and normal language at 24 months of age on measures of behavioral difficulties, social, and cognitive development. *Infant Mental Health Journal, 19*(1), 59-75.

Caspi, A., Elder, G.H., & Bem, D.J. (1987). Moving against the world: life-course patterns of explosive children. *Developmental Psychology, 23*, 308-313.

Cole, P.M., Usher, B.A., & Cargo, A.P. (1993). Cognitive risk and its association with risk for disruptive behavior disorder in preschoolers. *Journal of Clinical Child Psychology, 22*(2), 154-164.

*Cook, E.T., Greenberg, M.T., & Kusche, C.A. (1994). The relations between emotional understanding, intellectual functioning, and disruptive behavior problems in elementary-school-aged children. *Journal of Abnormal Psychology, 22*(2), 205-219.

Cooper, H. & Hedges, L. (Eds.). (1994). *Handbook of research synthesis*. New York: Russell Sage Foundation

Cornell, D.G. & Wilson, L. A. (1992). The VIQ < PIQ discrepancy in violent and nonviolent delinquents. *Journal of Clinical Psychology, 48*(2), 256-261.

Coy, K., Speltz, M.L., DeKlyen, M., & Jones, K. (2001). Social-cognitive processes in preschool boys with and without ODD, *Journal of Abnormal Child Psychology, 29*(2), 107-119.

Culberton, F.M., Feral, C.H., & Gabby, S. (1989). Pattern analysis of WISC-R profiles of delinquent boys. *Journal of Clinical Psychology, 45*(4), 651-660.

*Davis, A.D., Sanger, D.D., & Morris-Friehe, M. (1991). Language skills of delinquent and nondelinquent adolescent males. *Journal of Communication Disorders, 24*, 251-266.

Denno, D.W. (1986). Victim, offender, and situational characteristics of violent crime. *Journal of Criminal Law and Criminology, 77*(4), 1142-1158.

*Dery, M., Toupin, J., Pauze, R., Mercier, H., & Fortin, L. (1999). Neuropsychological characteristics of adolescents with conduct disorder: Association with attention-deficit-hyperactivity and aggression. *Journal of Abnormal Child Psychology, 27*(3), 225-236.

Dishion, T.J., Loeber, R., Stouthamer-Loeber, M., & Patterson, G.R. (1984). Skill deficits and male adolescent delinquency. *Journal of Abnormal Child Psychology, 12*(1), 37-54.

Essau, C.A. (2003). Epidemiology and risk factors. In C.A. Essau (Ed.), *Conduct and oppositional defiant disorders: Epidemiology, risk factors, and treatment* (pp. 33-59). New Jersey: Lawrence Erlbaum Associates.

Farrington, D.P. & Loeber, R. (2000). Epidemiology of juvenile violence. *Child and*

Adolescent Psychiatric Clinics of North America, 9(4), 733-749.

Fernald, P.S. & Wisser, R.E. (1967). Using WISC verbal-performance discrepancy to predict degree of acting out. *Journal of Clinical Psychology, 23*, 92-93.

Fitzhugh, K.B. (1973). Some neuropsychological features of delinquent subjects. *Perceptual and Motor Skills, 36*(2), 494.

Frazier, P.A., Tix, A.P., & Barron, K.E. (2004). Testing moderator and mediator effects in counseling psychology research. *Journal of Counseling Psychology, 51*(1), 115-134.

Frick, P.J., O'Brien, B.S., & Wootton, J.M. (1994). Psychopathology and conduct problems in children. *Journal of Abnormal Psychology, 103*(4), 700-707.

Frost, L.A., Moffitt, T.E., & McGee, R. (1989). Neuropsychological correlates of psychopathology in an unselected cohort of young adolescents. *Journal of Abnormal Psychology, 98*(3), 307-313.

*Giancola, P.R. & Mezzich, A.C. (2000). Executive cognitive functioning mediates the relation between language competence and antisocial behavior in conduct-disordered adolescent females. *Aggressive Behavior, 26*, 359-375.

*Golden, Z.L. & Golden, C.J. (2001). Do early onset conduct disordered adolescents perform like bran injured or normal adolescents on cognitive tests? *International Journal of Neuroscience, 111*, 109-121.

Gordon, N. (1991). The relationship between language and behavior. *Developmental Medicine and Child Neurology, 33*, 86-89.

Gorenstein, E.E. (1990). Neuropsychology of juvenile delinquents. *Forensic Reports, 3*, 15-48.

Grace, W.C. & Sweeney, M.E. (1986). Comparison of the P>V sign on the WISC-R and WAIS-R in delinquent males. *Journal of Clinical Psychology, 42*(1), 173-176.

Haynes, J.P. & Bensch, M. (1981). The P > V sign on the WISC-R and recidivism in delinquents. *Journal of Consulting and Clinical Psychology, 49*(3), 480-481.

Haynes, J.P. & Bensch, M. (1983). Female delinquent recidivism and the P>V sign on the WISC-R. *Journal of Clinical Psychology, 39*(1), 141-144.

Hays, J.R., Solway, K.S., & Schreiner, D. (1978). Intellectual characteristics of juvenile murderers versus status offenders. *Psychological Reports, 43*(1), 80-82.

Henning, J.J. & Levy, R.H. (1967). Verbal-performance IQ differences of white and

negro delinquents on the WISC and WAIS. *Journal of Clinical Psychology, 23*, 164-168.

Hirschi, T. & Hindelang, M.J. (1977). Intelligence and delinquency: A revisionist review. *American Sociological Review, 42*, 571-587.

Hogan, A.E. (1999). Cognitive functioning in children with ODD and CD. In *Handbook of Disruptive Behavior Disorders* (pp. 317-335).

Hubble, L.M. & Groff, M. (1981). Factor analysis of WISC-R scores of male delinquents referred for evaluation. *Journal of Consulting and Clinical Psychology, 49*(5), 738-739.

Hubble, L.M. & Groff, M. (1982). WISC-R verbal-performance IQ discrepancies among Quay-classified adolescent male delinquents. *Journal of Youth and Adolescence, 11(*6), 503-508).

Huesmann, L.R., Eron, L.D. (1984). Cognitive processes and the persistence of aggressive behavior. *Aggressive Behavior, 10*(3), 243-251.

*Humber, E. & Snow, P.C. (2001). The oral language skills of young offenders: A pilot investigation. *Psychiatry, Psychology, and Law, 8*(1), 1-11.

Hurwitz, I., Bibace, R.M., Wolff, P.H., & Rowbatham, B.M. (1972). Neurological function of normal boys, delinquent boys, and boys with learning problems. *Perceptual and Motor Skills, 35,* 387-394.

Jacob-Timm, S. & Hartshorne, T. (1998). Ethical issues in the education of pupils with disabilities under IDEA. In *Ethics and Law for School Psychologists (3rd Ed.).* New York: John Wiley & Sons, Inc.

Johnson-Reid, M., Williams, J.H., & Webster, D. (2001). Severe emotional disturbance and violent offending among incarcerated adolescents. *Social Work Research, 25*(4), 213-222.

*Karniski, W.M., Levine, M. D., Clarke, S., Palfrey, J.S., & Meltzer, L.J. (1982). A study of neurodevelopmental findings in early adolescent delinquents. *Journal of Adolescent Health Care, 3*151 -159.

Kaufman, J. (2001). Conduct disorder: overt aggression and conduct disorder: covert antisocial behavior. In *Characteristics of Emotional and Behavioral Disorders of Children and Youth (7th Ed.).* New Jersey: Prentice-Hall.

Kazdin, A.E. (2000). Treatments for aggressive and antisocial children. *Child and Adolescent Psychiatric Clinics of North America, 9*(4), 841-858.

*Kusche, C.A., Cook, E.T., & Greenberg, M. T. (1993). Neuropsychological and cognitive functioning in children with anxiety, externalizing, and comorbid psychopathology. *Journal of Clinical Child Psychology, 22*(2), 172-195.

Lahey, B.B. & Loeber, R. (1994). Framework for a developmental model of oppositional defiant disorder and conduct disorder. In: *Disruptive Behavior Disorders in Childhood*. D.K. Routh (Ed.), New York: Plenum Press.

Lahey, B.B., Loeber, R., Hart, E.L., et al. (1995). Four-year longitudinal study of conduct disorder in boys: Patterns and predictors of persistence. *Journal of Abnormal Psychology, 104*, 83-93.

Lewis, D.O., Shanok, S.S., Pincus, J.H., & Glaser, G.H. (1979). Violent juvenile delinquents: Psychiatric, neurological, and abuse factors. *Journal of the American Academy of Child Psychiatry, 18*, 307-319.

Lipsey, M.L. & Wilson, D.B. (2001). *Practical meta-analysis*: *Applied social research methods series* (vol.39). London: Sage Publications.

*Linz, T.D., Hooper, S.R., Hynd, G.W., Isaac, W., & Gibson, L.J. (1990). Frontal lobe functioning in conduct disordered juveniles: Preliminary findings. *Archives of Clinical Neuropsychology, 5*, 411-416.

Loeber, R. (1988). Natural histories of conduct problems, delinquency, and associated substance abuse: Evidence for developmental progressions. In B.B. Lahey & A.E. Kazdin (Eds.), *Advances in Clinical Psychology* (Vol. 11, pp.73-124). New York: Plemum Press.

Loeber, R. (1991). Antisocial behavior: more enduring than changeable? *American Academy of Child and Adolescent Psychiatry*, 30, 393-397.

Loeber, R., Burke, J.D., Lahey, B.B., Winters, A., & Zera, M. (2000). Oppositional defiant and conduct disorder: A review of the past ten years, part I. *Journal of the American Academy of Child and Adolescent Psychiatry, 39*(12), 1468-1484.

Loeber, R. & Coie, J. (2001). Continuities and discontinuities of development, with particular emphasis on emotional and cognitive components of disruptive behavior. In J. Hill & B. Maughan (Eds.), *Conduct disorders in childhood and adolescence* (pp.379-407). New York, New York: Cambridge University Press.

Loeber, R. & Dishion, T. (1983). Early predictors of male delinquency: A review. *Psychological Bulletin, 94*(1), 68-99.

Loeber, R., Green, S., Keenan, K., & Lahey, B. (1995). Which boys will fare worse? Early predictors of the onset of conduct disorder in a six-year longitudinal study. *Journal of the American Academy of Child Psychiatry, 34*(4), 499-509.

Loeber, R. & Stouthamer-Loeber, M. (1987). Prediction. In H.C. Quay (Ed.), *Handbook of juvenile delinquency* (pp. 325-382). Oxford, England: John Wiley & Sons.

Loeber, R. & Stouthamer-Loeber, M. (1998). Development of juvenile aggression and violence: some common misconceptions and controversies. *American Psychologist, 53*, 242-259.

Luria, A.R. (1961). *The role of speech in regulation in normal and abnormal behavior*. New York: Lippincott.

Luria, A.R. (1966). *Higher cortical functions in man*. Oxgford, England: Basic Books.

Lynam, D.R. & Henry, B. (2001). The role of neuropsychological deficits in conduct disorders. In J. Hill & B. Maughan (Eds.), *Conduct disorders in childhood and adolescence* (pp.235-263). New York, New York: Cambridge University Press.

Lynam, D.R., Moffitt, T., & Stouthamer-Loeber, M. (1993). Explaining the relation between IQ and delinquency: Class, race, test motivation, school failure, or self-control? *Journal of Abnormal Psychology, 102*(2), 187-196.

Mack, A.E., & Warr-Leeper, G.A. (1992). Language abilities in boys with chronic behavior disorders. *Language, Speech, and Hearing Services in Schools, 23*, 214-223.

Manne, S.H., Kandel, A., & Rosenthal, D. (1962). Differences between performance IQ and verbal IQ in a severely sociopathic population. *Journal of Clinical Psychology, 18*(1), 73-77.

Maskin, M.B. (1974). A comparison of graduate and recidivist WISC IQ scores in a delinquency program for girls. *Journal of Clinical Psychology, 30*(3), 319-320.

Mattison, R.E., Cantwell, D.P., & Baker, L. (1980). Behavior problems in children with speech and language retardations. *Child Psychiatry and Human Development, 10*(4), 246-257.

McMahon, R. & Wells, K. (1998). Conduct Problems. In E.J. Mash and R.A. Barkley (Eds.), *Treatment of Childhood Disorders (2nd Ed.).* (pp. 111-207). New York: The Guilford Press.

Meltzer, L.J., Roditi, B.N., & Fenton, T. (1986). Cognitive and learning profiles of delinquent and learning-disabled adolescents. *Adolescence, 21*, 581-591.

*Miniutti, A.M. (1991). Language deficiencies in inner-city children with learning and behavioral problems. *Language, Speech, and Hearing Services in Schools, 22*, 31-38.

Moffitt, T. E. (1988). Neuropsychology and self-reported early delinquency in an unselected birth cohort: A preliminary report from New Zealand. In *Biological Contributions to Crime Causation* (pp. 93-120).

Moffitt, T.E. (1990a). Juvenile delinquency and attention deficit disorder: Boys' developmental trajectories from age 3 to age 15. *Child Development, 61*, 893-910.

Moffitt, T.E. (1990b). The neuropsychology of delinquency: A critical review of theory and research. In *Crime and Justice: An Annual Review of Research, 12*, pp. 99-169.

Moffitt, T.E. (1993a). The neuropsychology of conduct disorder. *Development and Psychopathology, 5*, 135-151.

Moffitt, T.E. (1993b). Adolescence-limited and life-course persistent antisocial behavior: A developmental taxonomy. *Psychological Review, 100*(4), 674-701.

Moffitt, T.E., Gabrielli, W.F., Mednick, S.A., & Schulsinger, F. (1981). Socioeconomic status, IQ, and delinquency. *Journal of Abnormal Psychology, 90*(2), 152-156.

Moffitt, T.E., Lynam, D.R., and Silva, P.A. (1994). Neuropsychological tests predicting persistent male delinquency. *Criminology, 32*(2), 277-300.

Moffitt, T.E. & Silva, P.A. (1988a). Neuropsychological deficit and self-reported delinquency in an unselected birth cohort. *Journal of the American Academy of Child and Adolescent Psychiatry, 27*(2), 233-240.

*Moffitt, T. E. & Silva, P.A. (1988b). IQ and delinquency: A direct test of the differential detection hypothesis. *Journal of Abnormal Psychology, 97*(3), 330-333.

Moffitt, T.E. & Silva, P.A. (1988c). Self-reported delinquency, neuropsychological deficit, and history of attention deficit disorder. *Journal of Abnormal Child Psychology, 16*(5), 553-569.

Ollendick, T.H. (1979). Discrepancies between verbal and performance IQs and subtest scatter on the WISC for juvenile delinquents. *Psychological Reports, 45*(2), 563-568.

Olweus, D. (1979). Stability of aggressive reaction patterns in males: A review. *Psychological Bulletin, 86*(4), 852-875.

*Oosterlaan, J., Scheres, A., & Sergeant, J. (2005). Which executive functioning deficits are associated with ADHD, ODD/CD, and comorbid ADHD + ODD/CD? *Journal of Abnormal Child Psychology, 33*(1), 69-85.

Petee, T.A. & Walsh, A. (1987). Violent delinquency, race, and the Wechsler performance- verbal discrepancy. *Journal of Social Psychology, 127*(3), 353-354.

Prentice, N.M. & Kelly, F. J. (1963). Intelligence and delinquency: reconsideration. *Journal of Social Psychology, 60*, 327-337.

Psychological Corporation (2003). *Wechsler Intelligence Scale for Children-Fourth Edition Technical and Interpretive Manual*. The Psychological Corporation.

*Raine, A., Moffitt, T.E., Caspi, A., Loeber, R., Stouthamer-Loeber, M., & Lynam, D. (2005). Neurocognitive impairments in boys on the life-course persistent antisocial path. *Journal of Abnormal Psychology, 114*(1), 38-49.

Robbins, D.M., Beck, J.C., Pries, R., Jacobs, D., & Smith, C. (1983). Learning disability and neuropsychological impairment in adjudicated, unincarcerated male delinquents. *Journal of the American Academy of Child Psychiatry, 22*(1), 40-46.

Rosenthal, R. (1995). Writing meta-analytic reviews. *Psychological Bulletin, 118*(2), 183-192.

Sanger, D.D., Creswell, J.W., Dworak, J., & Schultz, L. (2000). Cultural analysis of communication behaviors among juveniles in a correctional facility. *Journal of Communication Disorders, 33*, 31-57.

Sanger, D.D., Hux, K., & Belau, D. (1997). Oral language skills of female juvenile delinquents. *American Journal of Speech-Language Pathology, 6*, 70-76.

Sanger, D.D., Hux, K., & Ritzman, M. (1999). Female juvenile delinquents' pragmatic awareness of conversational interactions. *Journal of Communication Disorders, 32*, 281-295.

Sanger, D.D., Moore-Brown, B., & Alt, E. (2000). Advancing the discussion on communication and violence. *Communication Disorders Quarterly, 22*(1), 43-48.

Sanger, D.D., Moore-Brown, B., Magnuson, G., & Svoboda, N. (2001). Prevalence of language problems among adolescent delinquents: A closer look. *Communication Disorders Quarterly, 23*(1), 17-26.

Sattler, J. (2001). *Assessment of children: Cognitive applications* (4th ed.). California: Jerome M. Sattler Publisher, Inc.

Schonfeld, I.S., Shaffer, D., O'Conner, P., & Portnoy, S. (1988). Conduct disorder and cognitive functioning: Testing three causal hypotheses. *Child Development, 59*, 993-1007.

Seashore, H.G. (1951). Differences between verbal and performance IQs on the WISC. *Journal of Clinical Psychology, 15*, 62-67.

Sharpe, D. (1997). Of apples and oranges, file drawers and garbage: Why validity issues in meta-analysis will not go away. *Clinical Psychology Review, 17*(8), 881-901.

Silva, P.A. (1990). The Dunedin multidisciplinary health and development study: A 15-year longitudinal study. *Paediatric and Perinatal Epidemiology, 4,* 76-107.

Silva, P.A., Williams, S., & McGee, R. (1987). A longitudinal study of children with developmental language delay at age 3: later intelligence, reading, and behavior problems. *Developmental Medicine and Child Neurology, 29,* 630-640.

Sobotowicz, W., Evans, J.R., & Laughlin, J. (1987). Neuropsychological functions and social support in delinquents and learning disabled. *The International Journal of Clinical Neuropsychology, 9*, 178-186.

Soeken, K.L. & Sripusanapan, A. (2003). Assessing publication bias in meta-analysis. *Nursing Research, 52*(1), 57-60.

Solway, K.S., Hays, J.R., & Roberts, T.K. (1975). Comparison of WISC profiles of alleged juvenile delinquents living at home versus those incarcerated. *Psychological Reports, 37*(2), 403-407.

Spellacy, F. (1977). Neuropsychological differences between violent and nonviolent adolescents. *Journal of Clinical Psychology, 33*, 966-969.

*Speltz, M.L., DeKlyen, M., Calderon, R., Greenberg, M.T., & Fisher, P.A. (1999). Neuropsychological characteristics and test behaviors of boys with early onset conduct problems. *Journal of Abnormal Psychology, 108*(2), 315-325.

Stattin, H. & Klackenberg-Larson, I. (1993). Early language and intelligence development and their relationship to future criminal behavior. *Journal of Abnormal Psychology, 102*(3), 369-378.

*Stellern, J., Marlowe, M., Jacobs, J., & Cossairt, A. (1985). Neuropsychological significance of right hemisphere cognitive mode in behavioral disorders. *Behavioral Disorders, February,* 113-124.

Stevenson, J., Richman, N., & Graham, P. (1985). Behavior problems and language abilities at 3 years and behavioral deviance at 8 years. *Journal of Child Psychology and Psychiatry, 26*(2), 215-230.

Stowe, R.M., Arnold, D.H., & Ortiz, C. (2000). Gender differences in the relationship of language development to disruptive behavior and peer relationships in preschoolers. *Journal of Applied Developmental Psychology, 20*(4), 521-536.

Tarter, R.E., Hegedus, A.M., Alterman, A.I., & Katz-Garris (1983). Cognitive capacities

of juvenile violent, nonviolent, and sexual offenders. *Journal of Nervous Mental Disorder, 171*, 564-576.

Tarter, R.E., Hegedus, A.M., Winsten, N.E., & Alterman, A.I. (1984). Neuropsychology, personality, and familial characteristic s of physically abused delinquents. *Journal of the American Academy of Child Psychiatry, 23*(6), 668-674.

Tarter, R.E., Hegedus, A.M., Winsten, N.E., & Alterman, A.I. (1985). Intellectual profiles and violent behavior in juvenile delinquents. *The Journal of Psychology, 119*(2), 125-128.

Teichner, G., Golden, C.J., Crum, T.A., Azrin, N.H., Donohue, B., & Van Hasselt, V.B. (2000). Identification of neuropsychological subtypes in a sample of delinquent adolescents. *Journal of Psychiatric Research, 34*, 129-132.

Tomblin, J.B., Zhang, X., Buckwalter, P., & Catts, H. (2000). The association of reading disability, behavior disorders, and language impairment among second-grade children. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 41*(4), 473-482.

Tramontana, M.G. & Hooper, S.R. (1987). Discriminating the presence and pattern of neuropsychological impairment in child psychiatric disorders. *International Journal of Clinical Neuropsychology, 9*(3), 111-119.

Tremblay, R.E., Pihl, R.O., Vitaro, F., & Dobkin, P.L. (1994). Predicting early onset of antisocial behavior from preschool behavior. *Archives of General Psychiatry*, 51, 732-739.

*Voorhees, J. (1981). Neuropsychological differences between juvenile delinquents and functional adolescents: A preliminary study. *Adolescence, 16*, 57-65.

Vygotsky, L.S., Hanfmann, E., & Vakar, G. (1962). *Thought and language*. New York: M.I. T. Press and Wiley.

Walsh, A. & Beyer, J.A (1986). Wechsler performance-verbal discrepancy and antisocial behavior. *The Journal of Social Psychology, 126*(3), 419-420.

Walsh, A., Petee, T.A., & Beyer, J.A. (1987). Intellectual imbalance and delinquency: Comparing high VIQ and high PIQ delinquents. *Criminal Justice and Behavior, 14*(3), 370-379.

Warr-Leeper, G., Wright, N.A., & Mack, A. (1994). Language disabilities of antisocial boys in residential treatment. *Behavioural Disorders, 19*(3), 159-169.

Webster-Stratton, C. (1993). Strategies for helping early school-aged children with

oppositional defiant and conduct disorders: the importance of home-school partnerships. *School Psychology Review, 22*(3), 437-457.

Werry, J.S., Elkind, G.S., & Reeves, J.C. (1987). Attention deficit, conduct, oppositional, and anxiety disorders in children: III. Laboratory differences. *Journal of Abnormal Child Psychology, 15*(3), 409-428.

West, D.J. & Farrington, D.P. (1973). Who *Becomes Delinquent?* London: Heinemann Educational Books.

White, J.L., Moffitt, T.E., & Silva, P.A. (1989). A prospective replication of the protective effects of IQ in subjects at high risk for juvenile delinquency. *Journal of Consulting and Clinical Psychology, 57*(6), 719-724.

Wickstrom-Kane, S. & Goldstein, H. (1999). Communication assessment and intervention to address challenging behaviors in toddlers. *Topics in Language Disorders, 19*(2), 70-89.

Wilson, W.J. & Herrnstein, R.J. (1985). *Crime and Human Nature*. New York: Simon and Schuster.

Wolff, P.H., Waber, D., Bauermeister, Cohen, C., & Ferber, R. (1982). The neuropsychological status of adolescent delinquent boys. *Journal of Child Psychology and Psychiatry, 23*(3), 267-279.

Wong, W. & Cornell, D.G. (1999). VIQ < PIQ discrepancy as a correlate of social problem solving and aggression in delinquent adolescent males. *Journal of Psychoeducational Assessment, 17*, 104-112.

Yeudall, L.T. (1980). A neuropsychological perspective of persistent juvenile delinquency and criminal behavior: Discussion. *Annals of the New York Academy of Sciences, 347*, 349-355.

*Yeudall, L.T., Fromm-Auch, D., & Davies, P. (1982). Neuropsychological impairment of persistent delinquency. *The Journal of Nervous and Mental Disease, 170*(5), 257-265.

Zincus, P.W. & Gottlieb, M.I. (1983). Patterns of auditory processing and articulation deficits in academically deficient juvenile delinquents. *Journal of Speech and Hearing Disorders, 48*(1), 36-40.