# SCHOOL-BASED SCREENING FOR MENTAL HEALTH DIFFICULTIES IN PRIMARY GRADE CHILDREN: PSYCHOMETRICS, INCREMENTAL VALIDITY, AND PATTERNS OF CO-OCCURRING DIFFICULTIES

by

Madison Lee Aitken

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of Applied Psychology and Human Development
Ontario Institute for Studies in Education
University of Toronto

**SCHOOL-BASED SCREENING FOR MENTAL HEALTH DIFFICULTIES IN PRIMARY GRADE CHILDREN: PSYCHOMETRICS, INCREMENTAL VALIDITY, AND PATTERNS OF CO-OCCURRING DIFFICULTIES**
**Doctor of Philosophy 2016**
**Madison Lee Aitken**
**Department of Applied Psychology and Human Development**
**University of Toronto**

**Abstract**

The demand for children's mental health services in Ontario exceeds available resources and, as a result, alternative approaches such as delivering mental health services in the school system may be needed. Within the school system, universal screening has been recommended as a method of identifying mental health difficulties early and beginning to provide intervention. If schools are to implement such screening programs, they must select appropriate screening measures and informants and be prepared to meet the needs of the children identified as at-risk. This process is complicated by high rates of co-occurrence among mental health difficulties, social and academic difficulties. The overall goal of this dissertation is to contribute to knowledge in the area of school-based mental health screening in terms of instruments, informants and the predominant patterns of co-occurring difficulties in elementary school children. Three studies will be presented, all of which involve the same community sample of 501 children in Grades 1 to 3.

In the first study, confirmatory factor analysis was used to examine the construct validity of the Strengths and Difficulties Questionnaire (SDQ), a potentially useful screening measure that has received little research attention in Canada. The results support the five-factor structure of parent and teacher SDQ ratings. Adequate internal consistency and inter-rater agreement were also found.

The second study used negative binomial regression to test the hypothesis that parent ratings of symptoms as well as parent and teacher ratings of impairment contribute to more accurate screening than teacher symptom ratings alone. The results suggest that the most useful combination of screening data includes teacher symptom and impairment ratings and parent impairment ratings.

The third study used latent class analysis to identify patterns of co-occurrence among some of the most common childhood difficulties: internalizing, externalizing, inattention/hyperactivity, social, and reading difficulties. The results indicate that a significant number of non-referred elementary school children experience difficulties in two or more important aspects of functioning.

Collectively, these findings provide an important foundation that may guide future research on and implementation of school-based mental health screening in terms of selecting measures and informants and anticipating potential patterns of need in young children.

# Table of Contents

# List of Tables

## List of Figures

# List of Abbreviations

ADHD = Attention-Deficit/Hyperactivity Disorder
APA = American Psychiatric Association
BASC = Behavior Assessment System for Children
BIC = Bayesian Information Criterion
BLRT = Bootstrap Likelihood Ratio Test
CBCL = Child Behavior Checklist
CFA = Confirmatory Factor Analysis
CFI = Comparative Fit Index
DIBELS = Dynamic Indicators of Basic Early Literacy Skills
DSM = Diagnostic and Statistical Manual of Mental Disorders
ICD = International Classification of Diseases
IRR = Incidence Rate Ratio
LCA = Latent Class Analysis
LPA = Latent Profile Analysis
RMSEA = Root Mean Square Error of Approximation
SDQ = Strengths and Difficulties Questionnaire
SPSS = Statistical Package for the Social Sciences
TLI = Tucker-Lewis Index
WHO = World Health Organization

**CHAPTER ONE**

**Introduction and Literature Review**

**1.1 Introduction**

The Mental Health Commission of Canada and the Ontario Government recently identified children and youth as a priority within the area of mental health services (Mental Health Commission of Canada, 2012; Ministry of Health and Long Term Care, 2011). This is especially important given the rates of mental health difficulties in Canadian children. The latest available review of Canadian studies found that between 13 and 18% of children and youth met criteria for a mental illness (Waddell, Offord, Shepherd, Hua, & McEwan, 2002). Similar findings have been reported more recently in an American sample, in which 12% of children and youth met diagnostic criteria for one or more disorders within a given year (Merikangas et al., 2010). These rates are even higher when sub-clinical levels of symptoms are considered, with as many as 23% of children experiencing mental health difficulties based on a large, nationally representative sample in the United States (Kataoka, Zhang, & Wells, 2002; Simon, Pastor, Reuben, Huang, & Goldstrom, 2015). Thus, the mental health of children in North America is a significant concern and is becoming more widely recognized as such.

The early identification of and intervention for mental health difficulties is an important component of preventive mental health approaches and schools may be an ideal setting for the delivery of these services (Ministry of Health and Long Term Care, 2011; School-Based Mental Health and Substance Abuse Consortium, 2013). Within the school system, universal screening for mental health difficulties has been recommended as a way to identify children who are experiencing difficulties and provide the services necessary to support them (Dowdy, Ritchey, & Kamphaus, 2010; School-Based Mental Health and Substance Abuse Consortium, 2013). However, mental health services in Canadian schools tend to focus on intervention for a small number of students with high needs rather than on prevention (School-Based Mental Health and

Substance Abuse Consortium, 2013). It is perhaps not surprising that there has been little uptake of screening approaches given the limited amount of research available to guide decisions about these programs (Dowdy & Kim, 2012), particularly in Canada. In addition, if schools elect to carry out screening, they must be prepared to meet the needs of the children identified as at-risk (R. Goodman, Ford, Simmons, Gatward, & Meltzer, 2000). This is complicated by high rates of co-occurrence among mental health difficulties as well as between mental health difficulties and social and academic problems (Angold, Costello, Farmer, Burns, & Erkanli, 1999; Dally, 2006; van Lier & Koot, 2010). The overall goal of this dissertation is to contribute to knowledge in the area of screening for mental health difficulties in Canadian children in ways that will inform decisions about appropriate measures and informants as well as the predominant patterns of co-occurring mental health, academic, and social difficulties identified through screening. Given the importance of early identification and intervention, the focus will be on the early elementary school years.

This dissertation is presented in portfolio format in five chapters. This chapter provides an overview of the literature on universal screening and related topics, including the psychometric properties of a promising screening measure, the utility of parent and teacher symptom and impairment ratings, and commonly co-occurring childhood difficulties and statistical approaches to studying them. The chapter concludes with the overall rationale, research questions, and hypotheses of the dissertation. Chapter 2 presents the first study, an examination of the psychometric properties the Strengths and Difficulties Questionnaire (SDQ; R. Goodman, 1997) in a sample of early elementary school students. The primary focus is a confirmatory factor analysis (CFA) of parent and teacher SDQ ratings. This manuscript was published in 2015 in the journal *Assessment for Effective Intervention*. Chapter 3 consists of a

study examining the predictive value of different types of information that can be gathered through screening, including parent and teacher ratings of symptoms and impairment.  In this way, it is possible to determine whether each piece of information contributes uniquely to the prediction of later functioning.  While Chapters 2 and 3 address measurement-related issues, Chapter 4 focuses on the results of screening in terms of patterns of co-occurring difficulties.  Co-occurring difficulties present unique statistical challenges and in this final study a relatively underutilized person-oriented approach is applied to complement previous multivariate studies.  Finally, Chapter 5 discusses the theoretical and practical significance of the findings and suggests directions for future research.  Because there is some overlap in the topics and the same sample of participants is used throughout the dissertation, there may be some repetition.

**1.2 Literature Review**

*1.21 Children's Mental Health*

The World Health Organization (WHO) defines mental health as "a state of well-being in which an individual realizes his or her own abilities, can cope with the normal stresses of life, can work productively and is able to make a contribution to his or her community" (WHO, 2014).  A subset of mental health disorders is common in children, including Attention-Deficit/Hyperactivity Disorder (ADHD), anxiety disorders, mood disorders, conduct problems, and oppositional defiance[1] (Angold, Costello, & Erkanli, 1999).  The focus of this dissertation is on sub-clinical difficulties and, as a result, the definition of terminology is important.  In young children, there is a great deal of overlap between anxiety and depression (Cole, Truglio, & Peeke, 1997; Wadsworth, Hudziak, Heath, & Achenbach, 2001) and the two are often referred to collectively as internalizing disorders (Kovacs & Devlin, 1998).  In keeping with the sub-clinical

---

[1] Substance use is also among the most common disorders of childhood and adolescence (Angold, Costello, & Erkanli, 1999); however, because this dissertation focuses on young children, substance use is not discussed here.

focus, I will use the term internalizing problems to refer to symptoms of anxiety and depression. Similarly, conduct problems and oppositional defiance are considered externalizing disorders (Jarrett & Ollendick, 2008) and their symptoms will be referred to here as externalizing problems. While ADHD is often included within the broader syndrome of externalizing disorders (e.g., Jenson, Harward, & Bowen, 2011), I will consider it separately given that it is classified as a neurodevelopmental disorder in the most recent edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) whereas conduct disorder and oppositional defiant disorder are classified under the heading of disruptive, impulse-control, and conduct disorders (APA, 2013). I will use the term inattention/hyperactivity to refer to the cluster of inattentive, hyperactive, and impulsive symptoms associated with the various subtypes of ADHD. Finally, I will refer to internalizing problems, externalizing problems, and inattention/hyperactivity collectively as mental health difficulties.

Mental health difficulties in childhood are associated with poor outcomes and functioning in both the short and long term. In particular, they negatively predict functioning in adolescence and early adulthood (Roza, Hofstra, van der Ende, & Verhulst, 2003) and are associated with a 35% reduction in subsequent family income in adulthood (Smith & Smith, 2010). Childhood mental health difficulties also have more proximal effects on individuals' functioning. For example, mental health difficulties in Grade 1 predict a range of negative outcomes later in elementary school, including greater symptom severity, poorer adaptive functioning, and lower academic achievement (Edelsohn, Ialongo, Werthamer-Larsson, Crockett, & Kellam, 1992; Ialongo, Edelsohn, Werthamer-Larsson, Crockett, & Kellam, 1995; Murphy et al., 2014). Therefore, early mental health difficulties are not transient and can have a lasting impact on individuals' functioning throughout childhood and into adulthood.

***Importance of Early Intervention.*** Early intervention is recognized as a central component of mental health services by both researchers and policymakers for several reasons (Bringewatt & Gershoff, 2010; Copeland et al., 2013; Kirby & Keon, 2006; Kratochwill, Albers, & Steele Shernoff, 2004; Mental Health Commission of Canada, 2012). First, early intervention not only provides relief in terms of symptoms and functioning during childhood but also potentially interrupts negative trajectories, leading to positive effects on long-term functioning (Copeland et al., 2013). Second, there can be significant economic benefits to preventing negative outcomes. For example, it is estimated that preventing externalizing behaviour in a single high-risk individual would result in a savings of 2.6 to 4.4 million dollars (Cohen & Piquero, 2009). Third, early intervention for mental health difficulties may be more effective than later intervention (Weare & Nind, 2011) given that effect sizes for psychotherapy are significantly larger for children than for adolescents (Weisz, Weiss, Alicke, & Klotz, 1987). Finally, early intervention and preventive approaches lead to statistically and clinically significant improvements in terms of current functioning and future outcomes (Dodge et al., 2015; Weare & Nind, 2011). Therefore, early intervention for mental health difficulties in childhood may have important individual, familial, and societal benefits in both the short- and long-term.

Despite the potential benefits of early intervention, many children with mental health needs do not receive specialized mental health services. An early Canadian study found that only 16% of children and youth who met diagnostic criteria for one or more mental health disorders had received specialized mental health services in the past six months (Offord et al., 1987). More recent data from the United States indicate that approximately half of children and youth with mental health needs have received some form of mental health services (Merikangas

et al., 2010; Simon et al., 2015). Information from Canadian organizations suggests that the need for mental health services for young people is growing. A recent analysis of hospital databases found that rates of mental health-related emergency department and inpatient stays among children and youth are increasing (Canadian Institute for Health Information, 2015) and the demand for community-based mental health services is rising by 10% annually (Children's Mental Health Ontario, 2015). Furthermore, Children's Mental Health Ontario (2015) reports that thousands of Ontario children are waiting an average of one year for psychological services. These figures indicate that meeting the mental health needs of Canadian children is an ongoing challenge.

*The Role of the School.* Researchers and policymakers have proposed that delivering mental health services within schools should be the primary strategy for increasing children's access to mental health services (Benningfield & Stephan, 2015; Bringewatt & Gershoff, 2010; Kirby & Keon, 2006; School-Based Mental Health and Substance Abuse Consortium, 2013). Several characteristics of schools make them especially well-suited to delivering mental health services. First, common barriers to mental health service use, such as stigma, transportation, and child care for siblings, are reduced or eliminated when these services are provided within the school, making mental health services more accessible to those least likely to access services elsewhere (Bringewatt & Gershoff, 2010; Stephan, Weist, Kataoka, Adelsheim, & Mills, 2007). Second, both targeted interventions and broader school-based interventions are possible, which may help to change the ecology of the school – one of the primary systems in which children function (Herman, Reinke, Parkin, Traylor, & Agarwal, 2009). Given that the school environment influences children's mental health (Herman et al., 2009), improvements at the school level may be a valuable intervention component. Third, interventions provided within the

school support generalization and have a high degree of ecological validity because they are delivered in the environment in which children function on a daily basis (Stephan et al., 2007). Finally, schools play a key role in reaching children with sub-clinical levels of symptoms, with 57% of these children initially being served through the school system (Farmer, Burns, Phillips, Angold, & Costello, 2003). As a result, schools are uniquely positioned to detect mental health difficulties and intervene early.

Currently, schools are one of the primary providers of mental health services for children, serving more children with mental health needs than other sectors such as specialized mental health or general medicine (Farmer et al., 2003; Offord et al., 1987; Simon et al., 2015). For example, the most recently available Canadian data indicate that 24% of children with mental health needs received services through the school system, compared to 16% who received specialized mental health services (Offord et al., 1987). Similarly, approximately 75% of children and youth in the United States who accessed mental health services at some point during their lifetime received services through the school system (Farmer et al., 2003). Schools are also the first point of entry into the mental health services system for the majority of children and it is common for schools to be the sole provider of services to children with mental health needs (Farmer et al., 2003; Simon et al., 2015). In addition, schools are frequently called upon to address the needs of children served by other sectors; for example, over half of the children who initially receive specialized mental health services subsequently receive services within the school system (Farmer et al., 2003). Thus, schools already dedicate a significant amount of resources to mental health; however, it is not clear that the services are being delivered in a way that maximizes their effectiveness.

Mental health services within the school system are generally individual and reactive, being provided only when a child experiences significant impairment (Albers, Glover, & Kratochwill, 2007; Dowdy, Ritchey, et al., 2010; Lane, Menzies, Oakes, & Kalberg, 2012; School-Based Mental Health and Substance Abuse Consortium, 2013). This approach is limited in its ability to improve children's mental health on a large scale because it is costly and does not change the broader school context in which the child is functioning (Dowdy, Ritchey, et al., 2010). Instead, a universal, preventive approach to mental health in the schools has been recommended (Dever, Dowdy, Raines, & Carnazzo, 2015; Dowdy, Ritchey, et al., 2010; Glover & Albers, 2007; Severson, Walker, Hope-Doolittle, Kratochwill, & Gresham, 2007). This would require a shift in focus from the individual level to the population and from intervention to prevention (Dowdy, Ritchey, et al., 2010); therefore, schools may need to change the way in which mental health services are delivered in order to be most effective.

### 1.2.2 Universal Screening

Screening programs are a potentially useful component of comprehensive school-based mental health services (Stephan, Sugai, Lever, & Connors, 2015). Universal screening for mental health difficulties is recognized as a best practice approach within the school system (Dever et al., 2015) and it has been recommended as a way of moving towards a prevention-focused, population-based approach to mental health (Dowdy, Ritchey, et al., 2010; Severson et al., 2007). Universal screening involves the systematic assessment of all children at regular intervals, not only those who have specific risk factors (Bruhn, Woods-Groves, & Huddle, 2014; Dowdy, Ritchey, et al., 2010). Students who are identified as at-risk then receive additional services which may include further assessment or intervention (Bruhn et al., 2014).

Several authors have proposed a three-tiered system as a framework for school-based mental health services (Dowdy, Ritchey, et al., 2010; Dvorsky, Girio-Herrera, & Owens, 2014; Kratochwill et al., 2004; Stephan et al., 2015). Tier 1, or universal interventions, are believed to be beneficial to all students, regardless of their risk status. These services are delivered to all students and include universal screening as well as school-wide initiatives such as teaching social and emotional skills (Dvorsky et al., 2014; Kratochwill et al., 2004; Stephan et al., 2015). It is estimated that this level of services will be sufficient to meet the needs of 80% of children (Stephan et al., 2015). Tier 2 interventions, sometimes referred to as selective interventions, are provided to a smaller number of students who have not responded to the Tier 1 intervention or who have multiple risk factors (Kratochwill et al., 2004). For example, students identified as at-risk through screening may receive Tier 2 services, such as individual or group intervention. Of the entire population of students, an additional 15% are expected to respond to Tier 2 interventions (Stephan et al., 2015). Tier 3, or indicated intervention, is more intensive than universal or selective interventions and is generally delivered to a small number of students who are experiencing significant difficulties and for whom the less intensive interventions have not been sufficient (Kratochwill et al., 2004). It is estimated that this will be approximately 5% of children (Stephan et al., 2015). Thus, universal screening fits with existing frameworks for mental health services within the school system (Dvorsky et al., 2014). In the following sections, universal screening is discussed in greater detail, including examples of current practices, potential benefits and criticisms, and unresolved implementation issues.

***Current Practices.*** Screening for mental health difficulties has lagged behind screening for academic difficulties (Bruhn et al., 2014; Cook, Volpe, & Livanis, 2010); however, there has recently been increased interest in and implementation of mental health screening (Lane et al.,

2012).  In a survey of school administrators in the United States, 12.6% said their school or district carried out school-wide mental health screening (Bruhn et al., 2014).  This is a substantial increase from rates reported a decade earlier, when it was estimated that approximately 2% of schools engaged in the practice (Romer & McIntosh, 2005).  The increased attention to mental health screening is encouraging given the importance of early identification and intervention; however, there is considerable room for improvement.  Below, two examples of screening and intervention programs are provided, one of which was carried out in the context of a large-scale research study whereas the other is more practically-oriented.

Dodge and colleagues (2015) conducted a randomized controlled trial of a screening and early intervention program for conduct problems.  In kindergarten, 9594 children from high-risk neighbourhoods were screened, first with teacher ratings and then with parent ratings for children with elevated teacher ratings.  Approximately 10% were identified as experiencing early conduct problems and received a 10-year comprehensive, manualized intervention that addressed various individual, environmental, and relational factors based on the child's needs.  Parents received behaviour management training, children received social-cognitive interventions as well as mentoring and tutoring, and changes to their classroom ecology were also made.  Positive effects were found at age 25 years in terms of psychiatric and substance abuse problems, crime, parenting self-efficacy and reduced use of physical punishment, and overall well-being.  For every eight children treated, a negative psychiatric outcome (internalizing, externalizing, or substance use) was prevented for one individual (Dodge et al., 2015).  This program was extremely comprehensive and is likely not feasible for widespread implementation; however, it illustrates that screening and early intervention can have lasting positive effects on development.

In contrast, Lane and colleagues (2011) present three simpler examples of school-based screening and intervention programs, one of which involved children from Kindergarten to Grade 4. In this example, a team within the school implemented a three-tier program. Tier 1 included teaching about appropriate behaviours on a school-wide basis and reinforcing these behaviours with a token economy system. In addition, teachers completed screening questionnaires at three time points that assessed internalizing, externalizing, and peer relationship difficulties. These data were used to determine the level of risk on a school-wide basis and to identify children for whom more intensive supports (Tier 2 or Tier 3) might be necessary. After one semester of the universal positive behaviour support intervention, rates of externalizing and internalizing problems decreased. Students who continued to have elevated scores on the screening measures and who had multiple risk factors then received further support, consisting of a functional assessment-based intervention (Lane et al., 2011). Because no statistical difference testing was carried out and there was no control group, it is not possible to determine the effectiveness of this multi-tier program; however, it is a useful illustration of universal screening and early intervention being implemented in a real-world context.

Further research is currently underway that examines the effectiveness of other screening and early intervention programs. For example, *Habilidades para la Vida* or Skills for Life is a government-run program for high-risk schools in Chile. Screening is carried out in elementary schools based on parent and teacher ratings and children who are at risk for mental health difficulties then attend preventive intervention groups. Preliminary findings suggest that the program is associated with significant academic and psychosocial improvement (Murphy et al., 2014). Thus, screening and early intervention are gaining momentum in both research and applied contexts and have demonstrated effectiveness and feasibility.

***Benefits and Criticisms.***   There are many potential benefits of universal screening for mental health difficulties, which were recently summarized by Dvorsky and colleagues (2014). First, screening is not subject to the same biases as traditional approaches to identifying mental health difficulties, such as teacher nomination (Dvorsky et al., 2014).  In particular, teachers' gender and whether difficulties are internalizing or externalizing in nature and at clinical or subclinical levels may influence their referral practices (Headley & Campbell, 2011; Loades & Mastroyannopoulou, 2010).  Moreover, behaviour rating scales identify more children as at risk than do teacher nominations (Dowdy, Doane, Eklund, & Dever, 2011; Eklund et al., 2009), suggesting that some children with mental health difficulties may be missed when teacher nominations are used, likely those with internalizing problems (Walker, 2010).  Second, universal screening may facilitate early intervention through the identification of difficulties before they progress to a clinical level (Albers et al., 2007; Dvorsky et al., 2014).  Third, the results of screening allow for the ongoing monitoring of students and provide a baseline to which future results can be compared (Dvorsky et al., 2014).  Finally, universal screening provides information on the needs of the entire population of students within a school or district (Lane et al., 2012), which may inform decisions about resource allocation and programming.  Therefore, screening may increase access to early intervention services and can also provide information that is useful at both the individual and population levels.

While there are many potential benefits of screening for mental health difficulties, concerns have also been raised about the practice.  In particular, identifying children as having mental health difficulties may be stigmatizing (Chafouleas, Kilgus, & Wallach, 2010) and this identification can lead to negative outcomes.  In a large study in the United Kingdom that screened for ADHD at kindergarten entry, children who were identified to school administrators

as exhibiting elevated ADHD symptoms were over twice as likely as children in a control group (not identified to administrators) to have elevated ADHD symptoms five years later (Sayal et al., 2010). This suggests that, in the absence of appropriate intervention, identification alone can have a detrimental effect on children's functioning in the long term, perhaps because it creates a negative expectation about the child (Sayal et al., 2010). A related concern is that screening measures attempt to strike a balance between false positives and false negatives and generally favour a higher rate of false positives in order to capture all children who may be at risk (Levitt, Saka, Romanelli, & Hoagwood, 2007). These false positives are problematic as they may cause undue stress to parents and children and create additional costs through the provision of unnecessary interventions (Feeney-Kettler, Kratochwill, Kaiser, Hemmeter, & Kettler, 2010). In addition, there are several ethical issues that must be considered (see Chafouleas et al., 2010 for a review), including informed consent, confidentiality, the use of psychometrically valid measures, the ability to provide appropriate services based on the needs identified, and fairness to all ethnic and cultural groups (Bruhn et al., 2014; Chafouleas et al., 2010; Kirby & Keon, 2006). These concerns and ethical issues are an important reminder that screening is only a small component of high quality mental health services for children and that it is not without risks. Encouragingly, however, a recent review-of-reviews found very few instances of adverse effects in school-based prevention and early intervention programs for mental health (Weare & Nind, 2011), suggesting that when early identification is tied to appropriate follow-up services, negative effects are unlikely.

An additional concern about mental health screening is that identifying a large number of children as at-risk will overwhelm school resources (Dever, Raines, & Barclay, 2012). While screening will initially result in more children being identified as requiring services, identifying

these difficulties early will mean that less intensive interventions are necessary to meet their needs (Dever et al., 2012).  Based on the limited evidence available, the number of children identified has generally been manageable given the available school or community resources (Levitt et al., 2007); however, further research in this area is necessary.

The cost of screening and early intervention programs is a major concern for schools (Glover & Albers, 2007) and only a handful of studies have addressed this issue.  The per-student cost of screening and follow-up assessment has been estimated at between $8.88 and $13.64 per student screened, depending on the number of positive screens (Kuo, Vander Stoep, McCauley, & Kernic, 2009).  When the intervention phase is considered, costs vary widely depending on the type of services provided, with annual estimates ranging from $674 to $1704 per student receiving counselling and/or psychopharmacological interventions to $5800 per student receiving a comprehensive intervention program (Chatterji, Caffray, Crowe, Freeman, & Jensen, 2004; Dodge et al., 2015).  Further research is needed on the cost of mental health screening programs as well as the potential cost-benefit of early detection and intervention programs (Dodge et al., 2015; Kuo et al., 2009); therefore, the figures above are provided for illustration purposes only.  In summary, universal screening may facilitate the early identification of mental health difficulties but ethical issues must be considered carefully and identification should be the first step in a continuum of mental health services.

***Implementation Questions.***  Schools wishing to implement a universal screening program for mental health difficulties face a number of decisions and often there is a lack of research to inform these choices.  For example, schools must choose measures that are valid and reliable, select informants, and decide how often to screen (Dowdy & Kim, 2012; Dowdy et al., 2014; Glover & Albers, 2007).  Much of the existing research on mental health screening has

focused on the psychometric properties of available measures and, as a result, several instruments are available that have demonstrated reliability and validity (for reviews, see Feeney-Kettler et al., 2010; Jenkins et al., 2014; Levitt et al., 2007).  There is also preliminary evidence that screening at multiple time points within a single year may be unnecessary (Dever et al., 2015; Dowdy et al., 2014), although further research with younger children is needed as this evidence is based on middle school and high school samples.

Much less is known about the best informants for screening purposes and whether it is necessary to gather ratings from multiple informants (Dowdy & Kim, 2012).  In addition, if screening is to be effective, schools must be prepared to address the needs of the students identified through the process (R. Goodman et al., 2000).  As a result, it is important for schools to have an understanding of the common patterns of mental health difficulties and how these are related to children's academic and social functioning.  In the sections that follow, I will review the psychometric properties of the SDQ (R. Goodman, 1997), a potentially useful tool for school-based mental health screening (Dowdy, Furlong, Eklund, Saeki, & Ritchey, 2010).  I will also summarize evidence related to the choice of parents[2] and/or teachers as informants and the type of information to collect from them.  Finally, I will discuss co-occurring childhood mental health, social, and academic difficulties and will describe a relatively novel approach to capturing the complexity of these patterns.

---

[2] In the majority of cases, respondent parents are mothers (Offord et al., 1996).  Mothers are more likely than fathers to complete behavioural rating questionnaires (Gizer et al., 2008; W. M. Reynolds, Anderson, & Bartell, 1985) and in the present dissertation 90% of respondent parents were mothers.  However, many studies do not distinguish between mothers and fathers as informants (e.g., Dirks et al., 2011; R. Goodman, 2001); for this reason, and because some respondent parents were fathers in the present data set, the term "parent ratings" will be used, unless a study being reviewed examined mother and father ratings separately.

### *1.2.3 The Strengths and Difficulties Questionnaire*

The SDQ (R. Goodman, 1997) is a broad-based behavioural rating scale that may be a useful screening tool for childhood mental health difficulties (Dowdy, Furlong, Eklund, Saeki, & Ritchey, 2010).  It consists of 25 items that are based on symptoms of disorders diagnosed in childhood and youth outlined in the DSM-IV (APA, 1994) and the International Classification of Diseases, 10th Edition (ICD-10; WHO, 1993), as well as on factor analyses (R. Goodman & Scott, 1999).  Each item is rated on a Likert-type scale and the items are divided equally into five subscales: emotional symptoms, conduct problems, hyperactivity-inattention, peer problems, and prosocial behaviour.  The first four subscales can be summed to generate a total difficulties score.  The SDQ can be completed by parents and teachers of 4 to 17 year-olds and by youth ages 11 to 16.  Because the focus of this dissertation is on early elementary school children, I will discuss only the parent and teacher versions of the SDQ, which are identical.  The SDQ is freely available at www.sdqinfo.org.

*Potential Utility.*  Several features of the SDQ make it especially useful as a measure for school-based mental health screening (for a review, see Stone et al., 2010).  First, it is freely available, which helps to reduce the overall cost of screening.  Second, it is relatively short, which reduces the demand on informants and, along with the inclusion of positively phrased items, may lead to higher completion rates (R. Goodman, 1994, 1997).  Third, the SDQ is available in more than 70 languages, making it especially useful for collecting information from parents from a range of linguistic backgrounds.  Fourth, the availability of identical parent and teacher versions makes it possible to gather information from multiple informants on the same behaviours.  Finally, the SDQ surveys a broad range of common childhood difficulties, including internalizing, externalizing, inattention/hyperactivity, and peer relationship problems.  Therefore,

information on several aspects of children's functioning can be gathered with a single, brief measure (Stone et al., 2010). These characteristics make the SDQ a practical choice as a screening measure for mental health difficulties.

*Reliability.* Numerous studies have examined the psychometric properties of various translations of the SDQ and, in general, the measure has demonstrated acceptable reliability based on evidence of internal consistency, inter-rater reliability, and test-retest reliability (for a review, see Stone, Otten, Engels, Vermulst, & Janssens, 2010). In terms of internal consistency, meta-analytic findings indicate that Cronbach's alpha values are acceptable for parent and teacher ratings on all of the SDQ subscales (Stone et al., 2010). Because alpha may underestimate reliability in measures with a small number of items (Graham, 2006), some researchers have used composite reliability as an indicator of internal consistency and have found slightly higher internal consistency values for the SDQ's subscales (e.g., Niclasen, Skovgaard, Andersen, Sømhovd, & Obel, 2013).

Inter-rater agreement between parent and teacher SDQ ratings is also satisfactory, particularly in comparison with other measures of child and adolescent mental health, ranging from $r = 0.24$ to 0.48 depending on the subscale. Parent-teacher agreement is at or above the meta-analytic mean of $r = 0.27$ reported for other measures (Achenbach, McConaughy, & Howell, 1987) for conduct problems (meta-analytic $r = 0.34$), hyperactivity-inattention (meta-analytic $r = 0.47$), peer problems (meta-analytic $r = 0.35$), and emotional symptoms (meta-analytic $r = 0.28$, although A. Goodman, Lamping, & Ploubidis, 2010, reported a correlation of $r = 0.24$) but slightly below the mean for prosocial behaviour (meta-analytic $r = 0.26$; Stone et al., 2010). In general, differences between parent and teacher SDQ ratings are comparable to those found for other measures.

Previous studies also support the test-retest reliability of the SDQ. Based on a review of six studies, mean test-retest correlations range from .65 to .71 for parent subscale ratings and from .72 to .85 for teacher subscale ratings (Stone et al., 2010). Stone and colleagues suggest that the slightly lower reliability of parent ratings may be because parents are more likely than teachers to notice differences in their child's mood and behaviour. In general, however, the test-retest reliability of the SDQ is adequate.

*Validity.* Evidence of the concurrent, predictive, and construct validity of the use of the SDQ results to make decisions about children has also been demonstrated with various samples of children and various versions of the SDQ. In terms of concurrent validity, the SDQ has shown moderate to high correlations with several broad and specific measures of mental health, including the Child Behavior Checklist (CBCL) and Teacher Report Form (Achenbach, 1991) and the parent versions of the Children's Depression Inventory and the Revised Children's Manifest Anxiety Scale (Stone et al., 2015; see Stone et al., 2010 for a review). Based on a meta-analysis, the correlation between the SDQ total score and CBCL and Teacher Report Form total scores is .76 (Stone et al., 2010). Correlations range from .52 to .71 between SDQ and similar CBCL subscales and from .57 to .79 between SDQ and similar Teacher Report Form subscales (Stone et al., 2010). In addition, the SDQ detects internalizing and externalizing behaviour as well as the CBCL and detects inattention/hyperactivity significantly better (R. Goodman & Scott, 1999). Collectively, these findings support the concurrent validity of the SDQ.

Little research has examined the predictive validity of the SDQ in terms of later functioning as measured by other indicators. However, one recent study of a sample of Dutch children reported that higher baseline parent SDQ ratings were associated with more parenting

challenges two years later and higher baseline teacher SDQ ratings predicted lower peer social preference ratings two years later (Stone et al., 2015). In conjunction with the evidence of concurrent validity, this suggests that the SDQ provides useful information about children's functioning that is associated with other indicators cross-sectionally and over time.

Finally, there is substantial evidence of the construct validity of the SDQ's subscales; however, much of the early research was exploratory (Niclasen et al., 2013) and confirmatory approaches have only been used recently. CFAs generally support the SDQ's five-factor structure (see Stone et al., 2010 for a review; Mieloo et al., 2012; Niclasen et al., 2013; Stone et al., 2015; van de Looij-Jansen, Goedhart, de Wilde, & Treffers, 2011), although a small number of studies have found that a three-factor solution, consisting of Internalization, Externalization, and Prosocial Problems, represents a better fit (e.g., Dickey & Blumberg, 2004; Koskelainen, Sourander, & Vauras, 2001). Likewise, some studies have found that bifactor models are needed in order to obtain a good fit (Caci, Morin, & Tran, 2015; Kóbor, Takács, & Urbán, 2013). Given these mixed findings, it is important to determine the adequacy of the SDQ's five-factor structure when using the measure in a new population of children.

### 1.2.4 First Research Goal

The SDQ is a potentially useful screening measure that has received little attention from Canadian researchers (see Woerner et al., 2004 for a review). Only a handful of published studies have reported using the SDQ as a screening measure in Canada (Huculak & McLennan, 2014; McLennan, Reckord, & Clarke, 2008) and no peer reviewed studies have examined the measure's psychometric properties in a Canadian sample. Providing evidence of the reliability and validity of the SDQ in a Canadian sample is an important prerequisite for increasing its use in research and practice. Therefore, the first goal of this dissertation is to confirm the factor

structure of the parent and teacher SDQ in a Canadian sample using CFA while also examining its reliability in terms of internal consistency and inter-rater agreement.

### 1.2.5 Incremental Validity

The use of a measure demonstrates evidence of incremental validity if its inclusion in an assessment battery results in better prediction of a criterion than that made based on the other tests in the battery (Sechrest, 1963). Confirming incremental validity is important in order to avoid the use of measures that do not contribute useful or unique information to an assessment (Sechrest, 1963). This is particularly relevant for school-based screening given the need for efficiency when carrying out such a broad assessment within a system whose resources are limited (Glover & Albers, 2007). Examining the incremental validity of information gathered when screening for mental health difficulties could therefore help practitioners to choose the most parsimonious set of measures.

Despite the potential value of incremental validity research to mental health screening, this topic has received little attention (for exceptions see Liu, Zhou, & Lackaff, 2013 who examined the incremental validity of measures designed to screen for autism spectrum disorder, developmental, and physical disabilities; and Owens et al., 2015 who examined the incremental validity of parent mental health ratings above and beyond academic screening). Moreover, few studies have addressed the issue of incremental validity as it pertains to the assessment of children and youth in general (De Los Reyes et al., 2015; Johnston & Murray, 2003). Therefore, further research is needed to ensure that mental health screening practices for children strike a balance between efficiency and accuracy.

***Potential Application to Screening.*** Based on current assessment practices for children's mental health, two parameters warrant examination in terms of incremental validity in a

screening context.  First, although a multi-informant approach is widely recommended in the psychological assessment of children (Hunsley & Mash, 2007; Johnston & Murray, 2003), it has been suggested that gathering information from multiple informants may not be necessary for screening (Dowdy & Kim, 2012).  This is in part because screening is intended to be an efficient process that leads to further services, including more comprehensive assessment (Dowdy & Kim, 2012; Glover & Albers, 2007; R. Goodman et al., 2000).  However, little peer-reviewed research exists to inform the choice of raters for screening purposes, leaving practitioners to rely on "clinical intuition and judgment" (Dowdy & Kim, 2012, p. 1).

A second aspect of measurement to consider is the type of information collected.  In addition to inquiring about the presence of symptoms, comprehensive mental health assessments consider impairment, which is generally required for the diagnosis of a mental disorder (American Psychiatric Association, 2013; Stringaris & Goodman, 2013); however, impairment ratings are not usually part of mental health screening (R. Goodman, 1999).  Therefore, in addition to examining the incremental validity of multi-informant ratings, it would be useful to examine the incremental validity of symptom and impairment ratings when screening for mental health difficulties.  Below, I will review existing research on the potential usefulness of multi-informant ratings and symptom and impairment ratings.

*Multiple Informants.*  School-based screening for mental health difficulties generally relies on teacher ratings (e.g., Lane et al., 2011; White, Connelly, Thompson, & Wilson, 2013).  While teacher ratings are a valid means of assessing children's mental health (Epkins, 1993; Hill, Coie, Lochman, & Greenberg, 2004; Sharp, Croudace, & Goodyer, 2005; Verhulst, Koot, & van der Ende, 1994), agreement between parent and teacher ratings is often low (Achenbach et al., 1987; De Los Reyes et al., 2015; Offord et al., 1996), suggesting that each informant provides

unique information.  Indeed, the use of both parent and teacher ratings contributes to more accurate prediction of concurrent mental health diagnoses as well as the longitudinal prediction of outcomes other than diagnosis (Bird, Gould, Rubio-Stipec, Staghezza, & Canino, 1991; Hill et al., 2004; Johnson, Hollis, Marlow, Simms, & Wolke, 2014; Lochman & The Conduct Problems Prevention Research Group, 1995; Power et al., 1998; Verhulst et al., 1994).  However, each of these studies has one or more characteristics that makes it difficult to generalize the results to a screening context, including the use of clinic-referred samples, a focus on diagnosis as an outcome, the use of lengthy, research-oriented measures, or a narrow focus on a particular type of disorder, often externalizing problems.  Moreover, a recent review of research on multiple informants in the assessment of children's mental health concluded that, due to methodological issues in previous studies, it is not possible to draw conclusions about the incremental validity of the multi-informant approach (De Los Reyes et al., 2015).  Therefore, further research is necessary to examine the incremental validity of parent and teacher ratings in general and, more specifically, in a context that more closely resembles screening.

*Symptom and Impairment Ratings.*  Current mental health screening practices generally involve gathering symptom ratings (R. Goodman, 1999); however, adding impairment ratings may be useful for several reasons.  First, considering impairment ratings may reduce the rate of false positives, particularly in community samples (DuPaul, Reid, An,astopoulos, & Power, 2014; Rapee, Bőgels, van der Sluis, Craske, & Ollendick, 2012; Rijlaarsdam et al., 2015).  Second, collecting impairment ratings may reduce the rate of false negatives by identifying children who are significantly impaired but who do not meet cut-off criteria based on symptoms (Angold, Costello, Farmer, et al., 1999; Gadow, Kaat, & Lecavalier, 2013; Huculak & McLennan, 2014; Wille, Bettge, Wittchen, & Ravens-Sieberer, 2008).  Finally, impairment

ratings are associated with meaningful indicators of functioning over time, including externalizing problems, mental health service use, and self-harm, even after controlling for baseline levels of symptoms (Pickles et al., 2001; Stringaris & Goodman, 2013). Thus, symptom and impairment ratings provide related but unique information (Gadow et al., 2013; Rapee et al., 2012) that may contribute to more accurate screening. While there is some evidence supporting the incremental validity of impairment ratings above and beyond symptom ratings (Pickles et al., 2001; Stringaris & Goodman, 2013), no peer-reviewed studies have examined the incremental validity of parent and teacher impairment ratings in a screening context. Thus, it is not clear whether each rater contributes unique information on impairment that is predictive of children's level of risk.

   *Analytic Approach.* Incremental validity can be demonstrated in numerous ways (see Hunsley & Meyer, 2003 for a review); however, it is important to use analytic approaches that consider both the unique and shared variance among the sources of data (Hunsley, 2003). As a result, hierarchical regression analyses predicting a target outcome, either longitudinally or cross-sectionally, have been recommended (Hunsley & Meyer, 2003). Hierarchical regression provides "a stringent test of validity because any shared variance between A and B that predicts the criterion is assigned to only A in such an analysis" (Hunsley & Meyer, 2003, p. 450). The choice of the criterion variable is also an important consideration (De Los Reyes et al., 2015; Hunsley, 2003). The criterion should be clinically relevant, reliable, and not based on the same information as any of the predictor variables (De Los Reyes et al., 2015; Hunsley, 2003). When this latter condition is not met, the association between the predictor and the criterion will be overestimated (Hunsley, 2003), which is referred to as "criterion contamination" (De Los Reyes et al., 2015, p. 15). Unfortunately, many existing studies of incremental validity in the

assessment of children's mental health suffer from this limitation (De Los Reyes et al., 2015). Therefore, studies of incremental validity must select criterion variables carefully so as to minimize the potential for contamination.

### 1.2.6 Second Research Goal

Little is known about the relative usefulness of multiple informants and symptom versus impairment ratings when screening for mental health difficulties in children. Therefore, the second goal of this dissertation is to determine whether parent and teacher impairment ratings and parent symptom ratings demonstrate incremental validity when considered alongside the usually-collected teacher symptom ratings. These findings will inform the selection of measures and informants for mental health screening within the school system.

### 1.2.7 Co-Occurring Childhood Difficulties

Instruments that can detect a broad range of common childhood mental health difficulties are ideal for school-based mental health screening (Levitt et al., 2007), especially because of the high rates of co-occurrence among childhood mental health difficulties (Angold, Costello, Farmer, et al., 1999; Baldwin & Dadds, 2008). In addition, mental health difficulties frequently co-occur with social and academic difficulties (Dally, 2006; Deater-Deckard, 2001; McGee, Partridge, Williams, & Silva, 1991; Rabiner & Coie, 2000; van Lier & Koot, 2010), suggesting that assessments should consider children's functioning across all three of these important domains. In the sections that follow, the co-occurrence of mental health, social, and academic difficulties will be discussed, including theoretical explanations for their co-occurrence, a review of evidence of their association, and a discussion of statistical approaches to the study of co-occurring difficulties. To reflect the sub-clinical focus of the present dissertation, the term "co-occurring" will be used throughout, instead of "comorbid."

Sample characteristics may influence the results of studies of co-occurring difficulties. For example, clinic-referred samples give misleading estimates of the strength of the associations among areas of difficulty because clinic-referred children exhibit more severe symptoms and impairment than children in community samples (Angold, Costello, & Erkanli, 1999; Wolff & Ollendick, 2006). Therefore, community samples are necessary in order to minimize methodological bias when studying co-occurring difficulties (Wolff & Ollendick, 2006). As a result, the review below focuses on studies using non-referred populations.

Participants' age should also be considered as it may relate to mental health functioning. In particular, some symptom presentations may be more common in one developmental period than another. For example, the risk of mood disorders increases greatly from childhood to adolescence (Roza et al., 2003) and conduct disorder is more common in adolescents than in children (Merikangas et al., 2010). There may also be age-related differences in patterns of comorbidity (Bird, Gould, & Staghezza, 1993; Waschbusch, 2002; Zahn-Waxler, Klimes-Dougan, & Slattery, 2000). Finally, disorders have different etiologies and contributing factors depending upon individuals' developmental level. For example, cognitions are more relevant to understanding mental health difficulties in older children than in younger children (Herman & Ostrander, 2007; Ostrander & Herman, 2006). Given these differences across the lifespan, results of studies examining comorbidity in one developmental period may not generalize to another. Because the focus of the present study is on early identification and school-based mental health services, the review below considers studies of elementary school children.

### 1.2.8 Theoretical Explanations

Several perspectives have been provided that attempt to account for the high rates of co-occurrence among childhood mental health difficulties (Angold, Costello, & Erkanli, 1999;

Lilienfeld, 2003).  Although it has been argued that methodological artifacts, such as shared

method variance, overlapping diagnostic criteria, or referral bias may account for the high rates

of co-occurring difficulties (Lilienfeld, 2003), this suggestion has received little empirical

support (Angold, Costello, & Erkanli, 1999; Wolff & Ollendick, 2006).  Instead, other

explanations, such as the presence of shared risk factors or difficulties in one area contributing to

difficulties in other areas, may be more useful (Lilienfeld, 2003; Wolff & Ollendick, 2006).

Below, I will review theories that fit with each of these perspectives and that are relevant to

understanding co-occurring difficulties in young children: the multiple deficit model

(Pennington, 2006) and the developmental cascades model (Masten et al., 2005).

*Multiple Deficit Model.*  The multiple deficit model was proposed by Pennington (2006)

in response to the inability of single deficit models to account for comorbidity among complex

behavioural disorders.  Single deficit models propose that one specific cognitive or etiological

factor produces a given disorder (Pennington, 2006).  In contrast, the multiple deficit model

proposes that complex behavioural disorders have multifactorial etiologies that involve

interactions between genetic or environmental risk and protective factors and no single

etiological factor can explain the occurrence of a disorder (Pennington, 2006); therefore,

comorbidity occurs when risk factors are shared among disorders.  Furthermore, because of these

shared risk factors, comorbidity is common among complex behavioural disorders (Pennington,

2006).  It is possible that similar processes underlie the co-occurrence of sub-clinical difficulties.

The multiple deficit model originated from the study of the comorbidity of reading

disability with ADHD or speech sound disorder (Pennington, 2006) and has most often been

applied to examining these co-occurring difficulties.  For example, in a twin study of ADHD and

reading disability in children and adolescents, slow processing speed was identified as the shared

risk factor associated with comorbid ADHD and reading disability (Willcutt et al., 2010).

Studies of other childhood disorders also provide support for the multiple deficit model. For

example, based on a meta-analysis of studies examining the neuropsychological underpinnings

of nine of the most common childhood disorders, several neuropsychological risk factors are

associated with more than one disorder (Willcutt, Sonuga-Barke, Nigg, & Sergeant, 2008).

Similarly, parental psychopathology, emotion regulation difficulties, and cognitive deficits or

distortions have been proposed as potential shared risk factors for conduct problems and

depression (Wolff & Ollendick, 2006). A recent study by Caspi and colleagues also provided

evidence of a general psychopathology factor in adults and demonstrated that those who scored

high overall on this factor demonstrated poorer self-control in childhood (Caspi et al., 2014).

Collectively, these studies support the view that complex behavioural disorders have

multifactorial etiologies and shared risk factors may contribute to comorbidity.

The primary application of the multiple deficit model has been to cognitive and

neuropsychological risk factors (e.g., McGrath et al., 2011; Willcutt et al., 2010) and, as a result,

it is useful to consider broader models within developmental psychopathology that may further

our understanding of the co-occurrence of common childhood difficulties. One such example is

the developmental cascades model.

***Developmental Cascades Model.*** The term "developmental cascades" refers to the

cumulative effects that interactions and transactions within a system have on development

(Masten & Cicchetti, 2010). These effects spread both within and across levels of the system

and can be unidirectional, bidirectional, or indirect (Masten & Cicchetti, 2010). Research on

developmental cascades has grown out of the field of developmental psychopathology and

encompasses associations between competence and mental health (Masten et al., 2005).

Accordingly, relationships between domains such as academic achievement, peer relationships, internalizing, externalizing, and inattention/hyperactivity have been examined (Masten et al., 2005; Murray-Close et al., 2010; van Lier & Koot, 2010). Developmental cascades can also include other aspects of the system, such as parenting practices or parents' mental health (Lewin-Bizan, Bowers, & Lerner, 2010; Pemberton et al., 2010); however, the present review will focus on associations among aspects of children's functioning.

Developmental cascades models have been tested in longitudinal studies spanning preschool to adulthood. For example, in a 20-year longitudinal study, externalizing problems in childhood were associated with academic difficulties in adolescence, which were in turn associated with internalizing problems in early adulthood (Masten et al., 2005). While this illustrates the developmental cascade effect, it does not provide information on potential interrelationships among aspects of functioning in childhood because measures were only collected once during childhood. In a study examining a shorter span of development, van Lier and Koot (2010) found evidence of a cascade from externalizing problems in Grade 1 to social difficulties over the next two years, which predicted both internalizing and externalizing symptoms in Grade 4. Therefore, cascade effects can be observed across the lifespan but also over shorter periods of time, such as the early elementary school years.

The studies cited above also demonstrate that the impact of mental health difficulties can extend into aspects of functioning such as academic and peer relationships, which can in turn further affect mental health. Therefore, common childhood difficulties may co-occur because difficulties in one area spread to others. However, positive developmental cascades are equally possible (e.g., Lewin-Bizan et al., 2010) and providing appropriate intervention may be a means of stopping maladaptive cascades or initiating a positive cascade, leading to benefits in other

aspects of functioning over time (Masten et al., 2005; Masten & Cicchetti, 2010). For example, the quality of the teacher-student relationship is associated with better child adaptation in terms of academic achievement, peer relationships, internalizing, and externalizing problems (Hughes, Cavell, & Willson, 2001; Pianta & Stuhlman, 2004; Wu, Hughes, & Kwok, 2010); as a result, improving teacher-student relationships may have effects in other important domains (Pianta & Stuhlman, 2004).

In summary, evidence of developmental cascades informs the present study in two ways: 1) by providing a potential explanation for the co-occurrence of difficulties across domains; and 2) by suggesting the need to assess competence domains, such as academics and peer relationships, alongside mental health in order to provide a comprehensive picture of children's functioning. Bringing together information from the multiple deficit model and developmental cascades model, co-occurring mental health, academic, and social difficulties are common in childhood and there are theoretical bases for this co-occurrence. In particular, comorbidity may arise because disorders share a common underlying deficit or risk factor and because difficulties affect children at a systemic level, leading to effects in other domains. This is in contrast to alternative explanations, whereby comorbidity is an artifact of chance, sampling bias, or rater bias (Pennington, Willcutt, & Rhee, 2005). In the following sections, evidence of the co-occurrence among mental health, academic, and peer relationship difficulties is reviewed.

### *1.2.9 Established Patterns of Co-Occurrence*

***Co-Occurrence among Mental Health Difficulties.*** Based on studies examining bivariate associations, there is evidence of greater-than-chance co-occurrence among the most common childhood mental health difficulties, namely inattention/hyperactivity, internalizing, and externalizing (Angold, Costello, & Erkanli, 1999; Bird et al., 1991; Costello, Mustillo,

Erkanli, Keeler, & Angold, 2003).  Meta-analytic findings indicate that children and adolescents who meet diagnostic criteria for ADHD are 10.7 times more likely than their peers to meet diagnostic criteria for an externalizing disorder and between 3.0 and 5.5 times more likely to meet criteria for an internalizing disorder (anxiety and depression, respectively; Angold, Costello, & Erkanli, 1999).  Similar evidence has been found in studies examining sub-clinical difficulties.  For example, the co-occurrence of inattention/hyperactivity and externalizing problems is 7.2 times more common than expected, suggesting that these difficulties co-occur systematically (Waschbusch, 2002).  Children who exhibit sub-clinical levels of inattention/hyperactivity also exhibit significantly more internalizing problems than their peers (Cho et al., 2009; Hong et al., 2014; Thorell & Rydell, 2008) and their level of internalizing problems is comparable to that of children who meet full diagnostic criteria for ADHD (Cho et al., 2009; Hong et al., 2014).  Therefore, inattention/hyperactivity, whether at clinical or sub-clinical levels, is associated with an increased likelihood of both internalizing and externalizing problems.

Internalizing and externalizing problems also co-occur at greater than chance rates (Wolff & Ollendick, 2006).  Angold and colleagues' meta-analysis indicated that the odds of children with externalizing disorders meeting diagnostic criteria for an internalizing disorder were between 3.1 and 6.6 times those of other children (for anxiety and depression, respectively; Angold, Costello, & Erkanli, 1999).  Similarly, in a longitudinal study of children followed from Kindergarten to Grade 7, mother-rated internalizing and externalizing behaviour in kindergarten were significantly correlated and increases in internalizing over time were related to increases in externalizing, both within mothers' ratings and across mother and teacher ratings (Keiley, Bates,

Dodge, & Pettit, 2000). Collectively, these studies indicate that there are high rates of co-occurrence among the most common childhood mental health difficulties.

      ***Co-Occurring Peer Relationship Problems.*** Peer relationships are an important component of children's overall development that have long been theorized to influence psychological functioning (e.g., Sullivan, 1953). It is therefore not surprising that peer relationship problems co-occur with a range of mental health difficulties. For the purposes of the present discussion, the term peer relationship problems refers to a range of social difficulties, including peer rejection, low social preference, having few friends, and feeling lonely. Several studies have reported concurrent associations between peer relationship problems and internalizing symptoms (Boivin, Poulin, & Vitaro, 1994; Keiley et al., 2000; Morrow, Hubbard, McAuliffe, Rubin, & Dearing, 2006). For example, in a sample of Grade 2 children, peer-rated social preference predicted significant variance in depressive symptoms, even after aggression had been accounted for (Morrow et al., 2006). Similarly, in a sample of Kindergarten children, those who were rejected by their peers exhibited significantly more internalizing problems than those who were not rejected (Keiley et al., 2000). Peer relationship problems are also common in children who exhibit externalizing behaviour (Deater-Deckard, 2001). Externalizing problems and popularity are negatively correlated in primary grade children (Hymel, Rubin, Rowden, & Lemare, 1990) and externalizing correlates positively with peer rejection, even in children as young as kindergarten-age (Keiley et al., 2000; Sturaro, Van Lier, Cuijpers, & Koot, 2011). Finally, children who exhibit inattentive/hyperactive behaviour are more likely than their peers to experience peer relationship problems (Cho et al., 2009) and inattention/hyperactivity also predicts lower peer social preference ratings (Bellanti & Bierman, 2000; Diamantopoulou, Henricsson, & Rydell, 2005; Diamantopoulou, Rydell, Thorell, & Bohlin, 2007). These studies

suggest that children with mental health difficulties are likely to also experience difficulties in some aspects of their peer relationships.

***Co-Occurring Reading Difficulties.*** Within the domain of academic functioning, reading is especially important because it is involved in almost all subject areas. Developing the ability to read has been referred to as "the traditional criterion of beginning school achievement" (Chapman, Tunmer, & Prochnow, 2000, p. 703) and beginning reading skills are a strong predictor of overall academic self-concept, even within the first two months of schooling (Chapman et al., 2000). Because reading ability is so salient to children and so important for their overall academic functioning, I will focus on difficulties in this area in the present review.

Reading difficulties co-occur with a wide range of mental health and social difficulties. Children who exhibit inattention/hyperactivity are especially at risk for reading difficulties, although most evidence suggests it is inattention and not hyperactivity that affects reading achievement (Dally, 2006; McGee et al., 1991; Rabiner & Coie, 2000). In addition, children who experience reading difficulties are more likely than their peers to exhibit internalizing problems (Margalit & Zak, 1984; Maughan & Carroll, 2006; Maughan, Rowe, Loeber, & Stouthamer-Loeber, 2003). The association between reading difficulties and depression is especially robust, holding even when socioeconomic status, inadequate parental supervision, physical punishment at home, conduct disorder, and inattention have been accounted for (Maughan et al., 2003). Significant associations between externalizing problems and reading difficulties have also been reported (Carroll, Maughan, Goodman, & Meltzer, 2005; Morgan, Farkas, Tufis, & Sperling, 2008; Willcutt & Pennington, 2000); however, it appears that this is primarily because children with externalizing problems also commonly exhibit elevated levels of inattention/hyperactivity (Carroll et al., 2005; Willcutt & Pennington, 2000). Finally, there is

some evidence of compromised peer relationships in children with reading difficulties, although much of this research is based on samples of older elementary school children (e.g., Bryan, 1976; Lopes, Cruz, & Rutherford, 2002).  In terms of evidence from younger samples, Grade 3 children with reading difficulties report feeling lonelier than other children (Al-Yagon & Margalit, 2006) and even children as young as 6 and 7 years are less popular than their peers if they experience reading difficulties (McMichael, 1980).  Of note, a recent study found that Grade 2 children with reading difficulties have significantly higher parent- and teacher-rated inattention/hyperactivity, internalizing, externalizing, and peer relationship problems on the SDQ than other children (Russell, Ryder, Norwich, & Ford, 2015).  Therefore, reading difficulties often co-occur with both peer relationship problems and mental health difficulties and the SDQ is able to detect these co-occurring difficulties in young children.

*Complex Interrelationships.*  The preceding review highlights the extensive evidence of pairwise co-occurrence among childhood mental health, social, and academic difficulties; however, this is an intentional over-simplification.  In actuality, there are complex interrelationships among these difficulties which may be bi-directional or mediated by the presence of other difficulties.  As an example of a bi-directional association, it appears that peer relationship problems are part of a vicious cycle that contributes to increases in hyperactivity and inattention in elementary school children (Tseng, Kawabata, Gau, & Crick, 2014).  Similarly, two studies that followed children from kindergarten to Grade 3 or 4 found a bi-directional association between peer rejection and externalizing problems over time, suggesting that each contributes to the growth and maintenance of the other (Sturaro et al., 2011; van Lier & Koot, 2010).  In terms of mediation, academic achievement accounts in part for the association between inattention and internalizing problems (depression) over the early elementary school

years (Herman, Lambert, Ialongo, & Ostrander, 2007). In addition, externalizing problems mediate the association between inattention/hyperactivity and peer relationship problems (Andrade & Tannock, 2013). As a result of these many interrelationships, it is necessary to control for or consider several other variables when examining associations among two or more areas of difficulty. In the following section, I will discuss statistical approaches to examining co-occurring difficulties, including the challenges associated with commonly used methods and a potential complementary approach that is currently underused.

### *1.2.10 Person-Oriented Analyses*

Statistical analyses can be classified as variable-oriented or person-oriented and variable-oriented approaches have predominated in the study of individual development and co-occurring difficulties (Bergman & Magnusson, 1997). Variable-oriented research is based on a conceptualization of individuals as "a summation of variables over time" (Bergman & Trost, 2006, p. 604). These studies use analyses such as correlation, regression, and structural equation modeling, which aim to demonstrate relationships among variables (Bergman & Magnusson, 1997). While these methods provide information about the prediction of outcomes and associations between the variables studied (B. O. Muthén & Muthén, 2000), it can be difficult to generalize their findings to describe individuals (Bergman & Magnusson, 1997).

In contrast, person-oriented approaches are based on the view that development is "a holistic, highly interactional, and individualized process" (Sterba & Bauer, 2010, p. 239). Individuals are understood as functioning entities made up of the interactions among behaviour, values, goals, biology, environment, and other factors (Bergman & Magnusson, 1997). Person-oriented approaches focus on patterns among individuals rather than patterns among variables (B. O. Muthén & Muthén, 2000). Thus, these methods can detect groups of individuals who are

alike in terms of the variables being studied (which signals similarities in their functioning) but different from individuals in other groups (Magnusson, 2003; B. O. Muthén & Muthén, 2000). Given the different underlying approaches and results, person-oriented analyses represent a useful complement to variable-oriented methods.

Person-oriented analyses are grounded in the following principles: 1) Aspects of development are individual-specific; 2) complex interactions occur in individual development, involving individual, environmental, temporal, and other factors; 3) individuals differ in how they change over time; 4) patterns of variables can be used to summarize development; 5) the variables comprised within these patterns cannot be meaningfully understood independent of one another; and 6) despite an infinite number of patterns being possible, a small number of patterns generally emerges (see Sterba & Bauer, 2010 for a review). Thus, person-oriented analyses identify patterns across variables that interact in complex ways to give rise to individual development, making them especially appropriate for the study of co-occurring difficulties.

***Advantages of Person-Oriented Analyses.*** Two other characteristics of person-oriented methods make them particularly useful for the study of co-occurring difficulties. First, considering interactions among multiple variables in variable-oriented analyses quickly becomes overwhelming (Cronbach, 1975); in contrast, person-oriented analyses can derive patterns across several variables, producing results that are more easily interpretable. Second, variable-oriented analyses often seek to identify unique predictors of a given outcome but this can obscure other predictors that may be related to the outcome in important ways (Haapasalo, Tremblay, Boulerice, & Vitaro, 2000). As Haapasalo and colleagues state, "the best predictor of an outcome is not necessarily the best or the only active component in a developmental process" (2000, p. 163). For example, they examined aggression, hyperactivity, inattention, anxiety, and

prosocial behaviour in kindergarten as predictors of later school placement.  In their variable-oriented analysis, inattention was among the best predictors of later school placement; however, when they examined patterns derived from a person-oriented analysis, even though children classified as "Bullies" (who were aggressive, nonprosocial, and hyperactive) did not exhibit high levels of inattention, they were almost equally at risk of being placed outside a regular classroom as were children who were classified as "Inattentive" (who exhibited inattentive and hyperactive behaviour; Haapasalo et al., 2000).  Thus, by allowing for a holistic analysis, person-oriented approaches increase our ability to detect complex interrelationships which may otherwise be missed.

   ***Types of Person-Oriented Analyses.***  Several types of person-oriented analyses are available and, among them, cluster analysis has been the most widely used (Reinke, Herman, Petras, & Ialongo, 2008; Wichstrøm & Berg-Nielsen, 2014).  Cluster analysis is a descriptive method, meaning that results are interpreted in the absence of objective fit indices (Sterba & Bauer, 2010).  As a result, some issues with the validity of the clustering solutions have been reported (Aitkin, Anderson, & Hinde, 1981).  In contrast, model-based analyses provide less biased results due to their allowance for measurement error and the availability of fit indices to determine the appropriateness of the grouping solution (Sterba & Bauer, 2010).  Latent class analysis (LCA) and latent profile analysis (LPA) are two examples of model-based person-oriented approaches and are differentiated by the type of observed variables used.  In LCA, the observed variables are categorical whereas LPA uses continuous variables.  In both cases, latent classes are identified based on the pattern of scores across the observed variables (B. O. Muthén & Muthén, 2000).

LCA and LPA generate conditional probabilities, which indicate the likelihood of a particular response given an individual's latent class membership (McCutcheon, 1987). Researchers can use these conditional probabilities to characterize and label the latent classes (Collins & Lanza, 2010). For example, if participants in Latent Class A have a 0.9 probability of exhibiting conduct problems, this means that 90% of the participants in that class exhibit conduct problems. In this example, if Latent Class A also had a 0.7 probability of internalizing problems, it might be labeled the Internalizing and Externalizing latent class. For the sake of clarity, the term "class" will be used throughout this dissertation to refer to latent classes of children derived from person-oriented analyses and the term "classroom" will be used to refer to children's school context.

***Previous Applications to Childhood Difficulties.*** A small number of studies have applied model-based person-oriented methods to the study of co-occurring childhood difficulties. In an early study, Tolan and Henry (1996) used LCA to examine patterns across the seven subscales of the Teacher Report Form (Achenbach, 1991) in Grade 1 to 6 children. They identified four latent classes: a normative class, a class with aggression plus one or more other problems, a class with aggression only, and a class with a low probability of aggression but moderate probabilities of other difficulties. This study provides initial evidence that LCA identifies patterns of difficulties that tend to co-occur in a sample of non-referred children; however, not all studies have identified classes characterized by co-occurring difficulties. For example, in a study of Chinese children living in poverty, analyses of self-reported depression, anxiety, self-esteem, and disruptive behaviour resulted in a three-class solution: normal, externalizing, and internalizing (Herman, Bi, Borden, & Reinke, 2012). This study provides evidence that pure externalizing and internalizing groups may also be present in community

samples.

Few person-oriented studies have examined patterns of academic and mental health difficulties in children.  In one such study, Reinke and colleagues (2008) used LCA to examine patterns across teacher-rated aggression, inattention, and oppositionality, and an objective measure of academic achievement in non-referred Grade 1 children.  They identified four latent classes of boys: one with academic and behaviour problems, one with academic problems only, one with behaviour problems only, and one with no problems.  The latent classes identified for girls were similar but no behaviour problems class emerged.  Comparable results were recently reported in a sample of Grade 3 children based on screening measures for academic and mental health difficulties.  In this latter study, three classes of comparable size were identified: one with minimal academic and mental health risk, one with mainly academic risk, and one at high risk academically and moderate risk in terms of mental health (King, Lembke, & Reinke, 2015). These studies suggest that some non-referred children experience only academic or only mental health difficulties, while others experience difficulties in both domains.  However, neither study examined internalizing problems separately, nor did they examine peer relationships, making it unclear how difficulties in these areas manifest alongside academic and externalizing problems. This is an important issue given that difficulties in these four areas tend to co-occur (Lopes et al., 2002; Masten et al., 2005; van Lier & Koot, 2010).

It is rare that studies, whether variable-oriented or person-oriented, simultaneously examine children's functioning across the domains of mental health, academics and peer relationships (for exceptions, see Baldwin & Dadds, 2008; Barriga et al., 2002; Drabick, Gadow, & Loney, 2008; Morgan, Farkas, Tufis, & Sperling, 2008).  Only one peer-reviewed study has used a model-based person-oriented analysis to examine patterns of difficulties across these three

areas. Valdez, Lambert, and Ialongo (2011) used LPA to identify subgroups of Grade 1 children based on self-reported depressive symptoms, peer-reported aggression and social acceptance, and an objective measure of academic achievement. Three classes were identified: one that was at risk behaviourally, academically, and in peer relationships, one that was at risk academically and in peer relationships, and one that was well-adjusted. Thus, when a wide range of common childhood difficulties is examined, classes characterized by co-occurring difficulties emerge in the absence of classes characterized by single difficulties. This suggests that there is a relatively low likelihood of a child experiencing difficulties in an isolated domain even in non-referred samples. However, the authors noted that future analyses should include inattention (Valdez et al., 2011), which is especially relevant given the high rates of co-occurrence among inattention/hyperactivity and difficulties in other areas (e.g., Cho et al., 2009).

### 1.2.11 Third Research Goal

Existing person-oriented studies provide further evidence that internalizing, externalizing, inattention/hyperactivity, peer relationship problems, and academic difficulties tend to co-occur, even in community samples of children. However, to my knowledge, no studies have used model-based person-oriented analyses to examine patterns across these five important domains of children's functioning. Moreover, it is rare that analyses of any type examine all five of these domains using multivariate methods. Given the many variables involved and the demonstrated interrelationships among them, a person-oriented analysis is needed in order to reveal patterns without obscuring important relationships due to shared variance. An analysis of this type that uses data from screening measures may be especially useful as it would provide information on the patterns of needs likely to be identified when screening for mental health, social, and academic difficulties. The goal of the third study was to apply LCA to screening ratings of

internalizing problems, externalizing problems, inattention/hyperactivity, peer relationship problems, and reading difficulties to determine what patterns of co-occurrence are present in a community sample of early elementary school children.

### *1.2.12 Canadian Context*

The Ontario Child Health Study was carried out in 1983 and included parent, teacher, and self-report ratings of the mental health functioning of children and youth ages 4 to 16. Thirty years later, the study is currently being repeated in order to provide updated information on children's mental health in Ontario (Ontario Child Health Study, 2014). A similar study was also conducted in Quebec in 1992 (Breton et al., 1999). More recent information is available on the mental health of adolescents in Ontario, among whom 22% report having seen a mental health professional within the past year and 26% report experiencing psychological distress in the past month (Boak et al., 2014). Therefore, a significant proportion of young Canadians are struggling with mental health difficulties; many of these children and adolescents may have benefitted from early intervention services.

There is a consensus among recent Canadian policy documents that children's mental health should be a key priority and that more school-based mental health services are essential (Ministry of Health and Long Term Care, 2011; School-Based Mental Health and Substance Abuse Consortium, 2013). However, most of the school-based mental health programming currently in place focuses on adolescents rather than younger children (School-Based Mental Health and Substance Abuse Consortium, 2013). In addition, although Canadian policy documents have identified screening as a useful approach to preventing mental health difficulties (Ministry of Health and Long Term Care, 2011; School-Based Mental Health and Substance Abuse Consortium, 2013), a search of a database of current school-based mental health practices

in Canada ([www.excellenceforchildrenandyouth.ca/sites/scandb](www.excellenceforchildrenandyouth.ca/sites/scandb)) locates only two screening

programs, both of which target adolescents (although see Huculak & McLennan, 2014 and

McLennan, Reckord, & Clarke, 2008 for examples of using the SDQ to screen clinic-referred

children in Alberta).  This is perhaps not surprising given the dearth of research on mental health

screening in Canada.  Further high-quality research within a Canadian context is a necessary

prerequisite for the implementation of effective screening programs to identify mental health

difficulties in young children and begin providing appropriate interventions.  It is this need that

has informed the development of the three studies that make up this dissertation.

**1.3 The Current Dissertation**

The preceding review makes a case for the early identification of children's mental health

difficulties through school-based screening and highlights the need to consider mental health

alongside social and academic functioning.  A number of screening measures have demonstrated

reliability and validity based on international studies; however, no peer-reviewed studies have

examined the psychometric properties of mental health screening measures in a Canadian

sample.  In addition, there is a lack of research to inform important decisions about screening,

such as the choice of informants and the type of information gathered.  Given the high rates of

co-occurrence among mental health, academic, and peer relationship problems in childhood, it is

also important to determine the most prevalent patterns of difficulties in young children in order

to inform intervention planning.  The current dissertation will address these gaps in existing

research in three studies, which are outlined below.

*1.3.1 Research Questions and Hypotheses*

The overall objective of the first study (Chapter 2) was to examine the psychometric

properties of the SDQ, a potentially useful screening tool for mental health difficulties, in a

sample of young Canadian children.  It addresses the following research questions: 1) Is the SDQ a reliable measure in terms of internal consistency and inter-rater (parent and teacher) agreement?; and 2) Is the five-factor structure of the SDQ valid based on evidence from a CFA? Given that there is substantial evidence of the reliability and validity of the SDQ in other samples, it was anticipated that the measure would demonstrate acceptable psychometric properties in the present Canadian sample.

The second study (Chapter 3) sought to provide information on incremental validity in order to guide decisions about the information gathered when screening for mental health difficulties.  Specifically, it addresses the following research question: Do parent and teacher ratings of children's symptoms and impairment each contribute unique information to the prediction of children's functioning one year later?  Given the limited agreement between parent and teacher ratings, it was hypothesized that all four predictors (teacher-rated symptoms and impairment, parent-rated symptoms and impairment) would uniquely predict later functioning.

Finally, the third study used LCA to determine what patterns of co-occurrence exist among the most common childhood difficulties in a community sample of young children (Chapter 4).  The third study addresses the following research questions: 1) What patterns of co-occurring internalizing, externalizing, inattention/hyperactivity, peer relationship problems, and reading difficulties are present in a sample of non-referred children based on parent and teacher ratings?; and 2) Are these patterns meaningfully associated with differences in functioning over time?  Although this study was largely exploratory, it was expected that most children would fall within a low-risk latent class and that the remainder would be distributed among latent classes characterized by co-occurring difficulties.  Inattention/hyperactivity was expected to co-occur with other difficulties in one or more latent classes.

# CHAPTER TWO


**Factor Structure of the Strengths and Difficulties Questionnaire**

**2.1 Abstract**

The Strengths and Difficulties Questionnaire (SDQ) is a 25-item screening measure for emotional and behavioural problems in children and adolescents ages 4 to 16. Structural equation modeling was used to test the five-factor structure of teacher and parent ratings on the British version of the SDQ in a community sample of 501 Canadian children ages 6 to 9 years. The five-factor model fit the data well for both teacher and parent ratings. In addition, the SDQ demonstrated acceptable reliability based on internal consistency coefficients and inter-rater agreement. The findings provide preliminary support for the reliability and validity of the British version of the parent and teacher SDQ in a sample of early elementary school Canadian children.

**2.2 Introduction**

In Canada, only 16% of children with mental health needs receive services through a specialized mental health agency, but many more (24%) receive services through the school system (Offord et al., 1987). It has been suggested that delivering mental health services within the school is a means of reaching children and youth who are less likely to access services elsewhere (Bringewatt & Gershoff, 2010). Therefore, schools represent an important site for the identification of and intervention for mental health difficulties.

Within the school system, teacher nominations are frequently used to identify children in need of services; however, concerns have been raised about teacher referrals as a means of identifying children with mental health needs (Dvorsky et al., 2014). For example, teacher gender may influence their referral practices, with female teachers being more likely than male teachers to refer children with anxiety symptoms to a guidance counselor (Headley & Campbell, 2011). In addition, teachers view externalizing symptoms as more concerning than internalizing symptoms (Loades & Mastroyannopoulou, 2010), which may influence the threshold at which they perceive additional support to be necessary for the child. While there is evidence that teachers can recognize emotional and behavioural difficulties that fall within the clinical range, they may be less accurate at identifying difficulties falling within the sub-clinical range (Headley & Campbell, 2011; Loades & Mastroyannopoulou, 2010). Indeed, the use of a structured behaviour rating scale identifies more children as at risk for mental health difficulties than do teacher nominations (Dowdy et al., 2013). As a result, behaviour rating scales may be superior to teacher nominations if early identification and prevention of mental health difficulties are the goals (Dvorsky et al., 2014).

While teachers can provide important information on children's functioning (Epkins,

1993), considering parent and teacher ratings may lead to more accurate identification of children at risk for mental health difficulties (R. Goodman et al., 2000). Correlations between parent and teacher ratings are often relatively low (Achenbach et al., 1987), which may be due in part to actual differences in children's behaviour across settings (De Los Reyes, Henry, Tolan, & Wakschlag, 2009). In addition, parent and teacher ratings of mental health independently predict long-term functioning into adulthood (Dirks, Boyle, & Georgiades, 2011). Therefore, having both parents and teachers complete a valid and reliable rating scale appears to be an appropriate means of identifying children at risk for mental health difficulties.

### 2.2.1 The Strengths and Difficulties Questionnaire

The SDQ (R. Goodman, 1997) is one rating scale that may be useful in screening for social, emotional, behavioural, and attentional difficulties (Dowdy, Furlong, et al., 2010). The SDQ is a behavioural screening measure that can be completed by parents and teachers of 4 to 16 year-olds. It consists of 25 items based on symptoms of disorders diagnosed in childhood, as outlined in the DSM-IV (APA, 1994) and the ICD-10 (R. Goodman & Scott, 1999; WHO, 1993). Items are rated on a Likert-type scale (not true, somewhat true, certainly true) and are divided equally into five subscales: emotional symptoms, conduct problems, hyperactivity-inattention, peer problems, and prosocial behaviour. The first four subscales can be summed to generate a total difficulties score.

Several features of the SDQ make it a potentially useful measure for screening within the school system. First, the SDQ was designed to be user-friendly due to its length and the inclusion of positively phrased items (R. Goodman, 1994, 1997). The SDQ was also designed to be applicable across childhood and youth (R. Goodman, 1994) and parallel teacher- and parent-report versions are available, allowing for the use of multiple informants. The SDQ has been

translated into over 70 languages, making it a potentially useful means of gathering information from parents who may not have sufficient language proficiency to complete questionnaires in English.  Finally, the SDQ is freely available ([www.sdqinfo.org](www.sdqinfo.org)) and is therefore a cost-effective means of gathering ratings from parents and teachers.

*The SDQ as a Screening Instrument.*  The SDQ subscale scores and the total difficulties score can be classified as normal, borderline, or abnormal.  Cut-off scores corresponding to these classifications have been chosen so that in a community sample approximately 10% of children and adolescents score within the abnormal range and 10% score within the borderline range (R. Goodman, 1997).  In an epidemiological sample of British children ages 5 to 15 ($n = 10,438$), children were 15 times more likely to be assigned a DSM-IV diagnosis if they fell within the abnormal range (versus the borderline or normal ranges) for any of the parent- or teacher-rated scores (R. Goodman, 2001).  Classifying children and youth in this way resulted in specificity values for the total difficulties and subscale scores ranging from 91% to 94% for parent ratings and from 91% to 95% for teacher ratings.  Sensitivity values were lower and more variable, ranging from 25% (prosocial behaviour) to 74% (hyperactivity-inattention) for parent ratings and from 25% (peer problems) to 68% (hyperactivity-inattention) for teacher ratings.  Results suggested that the scoring cut-offs perform well in terms of reducing false negatives but they may also lead to false positives (R. Goodman, 2001).  R. Goodman further suggested that combining information across raters and considering the impact of the child's symptoms on their functioning may lead to more accurate classification.

The SDQ may tend to identify false positives, particularly in a community sample (see Stone et al., 2010 for a review); however, this "overinclusiveness" (R. Goodman, 2001, p. 1343) may be acceptable in a screening context when the goal is the early identification of mental

health difficulties (R. Goodman, 2001). Some school districts have begun using the SDQ to assess children's social, emotional, and behavioural functioning or to screen for mental health difficulties. Examples of this include using the SDQ to inform transition planning for preschool children (White et al., 2013) or using the SDQ in conjunction with another behaviour rating scale to identify students who may need more targeted intervention (Lane et al., 2011).

***Psychometric Properties of the Parent and Teacher SDQ.*** Evidence of the reliability and validity of the parent and teacher SDQ is accumulating based on various translations of the measure across a range of countries and populations (e.g., community samples vs. clinic-referred samples; see Stone et al., 2010 for a meta-analysis; Woerner et al., 2004 for a review). However, no peer reviewed studies have examined the psychometric properties of the SDQ in a Canadian sample. The present review focuses on studies examining the internal consistency, inter-rater reliability, and factor structure of the parent and teacher report forms of the British SDQ.

*Internal Consistency.* Cronbach's alpha has been used as an indicator of internal consistency for parent- and teacher-rated SDQ subscales in studies using nationally representative samples of children in England ($n = 9998$ for parent ratings, $n = 7313$ for teacher ratings; children ages 5-15 years; R. Goodman, 2001), Ireland ($n = 8568$ 9-year-old children with parent ratings; McCrory & Layte, 2012), and England, Scotland, and Wales ($n = 18,222$ with parent ratings, $n = 14,263$ with teacher ratings; A. Goodman et al., 2010). Cronbach's alpha values ranging from 0.52 to 0.77 for parent ratings and from 0.69 to 0.88 for teacher ratings have been reported. Across studies, the peer problems subscale has had the lowest internal consistency coefficients and the hyperactivity-inattention subscale the highest for both parent and teacher ratings (A. Goodman et al., 2010; R. Goodman, 2001; McCrory & Layte, 2012). Because alpha can underestimate reliability, particularly for scales with a small number of items

(Graham, 2006), these values may not accurately represent the internal consistency of the SDQ subscales. It has been suggested that alternative measures such as composite reliability may better represent the subscales' internal consistency (Niclasen et al., 2012). However, studies that have used composite reliability to measure the internal consistency of the British version of the SDQ have not been located.

*Inter-Rater Reliability.* Previous studies have compared inter-rater correlations for the SDQ to the meta-analytic mean of $r = .27$ for inter-rater agreement across measures (Achenbach et al., 1987; R. Goodman, 2001). In community samples from England ($n = 7313$ children ages 5 to 15 years; R. Goodman, 2001) and a nationally-representative sample of children ages 5 to 15 years from England, Scotland, and Wales ($n = 14,139$; A. Goodman et al., 2010), correlations between parent and teacher SDQ ratings have ranged from .24 to .48. Inter-rater correlations were at or above the meta-analytic mean for all subscales except prosocial behaviour ($r = .24$ or .25; A. Goodman et al., 2010; R. Goodman, 2001) and emotional symptoms ($r = .24$; A. Goodman et al., 2010). These results suggest that the SDQ performs comparably to other measures of children's social, emotional, and behavioural functioning in terms of parent-teacher agreement.

*Factor Structure.* Both exploratory factor analysis and CFA have been applied to examine the factor structure of the SDQ. An exploratory factor analysis of the SDQ based on an epidemiological sample of 5 to 15 year-old British children ($n = 9998$ for parent ratings, $n = 7313$ for teacher ratings) suggested that the five subscales adequately represented both parent and teacher ratings (R. Goodman, 2001). However, subsequent CFA studies have not provided such definitive support for a five-factor model (A. Goodman et al., 2010; McCrory & Layte, 2012). A. Goodman and colleagues (2010) used CFA to examine the factor structure of the SDQ

in a nationally-representative sample of children ages 5 to 15 years from England, Scotland, and Wales ($n = 18,222$ with parent ratings, $n = 14,263$ with teacher ratings). They found that a five-factor model with some items allowed to correlate within-factors showed an acceptable (but not good) fit for parent and teacher ratings. Similarly, in McCrory and Layte's (2012) sample of 9-year-old children in Ireland, the five factor structure only met criteria for an adequately fitting model when an additional minor factor tapping positive construal was added. However, in this latter study the authors did not attempt to improve the fit of the five-factor model by allowing items to correlate within factors.

### 2.2.2 Current Study

CFA studies to date provide only partial support for the five-factor structure of the parent and teacher SDQ. As a result, when using the SDQ in a new population, such as the Canadian sample used in the present study, it is important to determine the adequacy of its five factor structure. The goals of the present study were to use CFA to confirm the five-factor structure of the British version of the parent and teacher SDQ in a community sample of Canadian children, as well as to examine internal consistency and inter-rater agreement for the SDQ's subscales.

## 2.3 Method

### 2.3.1 Procedure

This study is part of a larger longitudinal study of behavioural and cognitive measures of attention and academic outcomes in elementary school children. The reader is directed to Normand, Flora, Toplak, and Tannock (2012) and Andrade and Tannock (2013) for detailed descriptions of the sample and procedures. Participants in Grades 1, 2, and 3 were recruited from one rural and one suburban district school board in Southern Ontario, Canada. School principals were invited to contact the researchers if they were interested in having their school

participate. Information sessions were then provided for teachers of Grades 1 to 3 in participating schools. Consenting teachers were given research packages containing a cover letter, study information, consent forms, and questionnaires, which they gave to parents (Normand et al., 2012). Teachers were also given questionnaires to complete for each participating child. Data were collected in November to ensure that teachers had sufficient time to observe and interact with the children before completing the rating scales (Normand et al., 2012). Data used in the present study are from the first measurement time point in a 2-year prospective study.

### 2.3.2 Sample

Participants were 501 children (247 boys, 254 girls) ranging in age from 6 to 9 years (mean age = 7.5 years). Eligibility criteria included: (a) being in a mainstream classroom in either English or French (25% were in French immersion); (b) no sensory or physical impairments that would preclude the child from completing the tasks or hearing the instructions; and (c) written parental and teacher informed consent and verbal child assent (Normand et al., 2012). Seven schools took part in the study, representing 20% of the schools in the board. All Grade 1 to 3 teachers ($n = 51$) in the participating schools consented to take part in the study (see Normand et al., 2012 for a detailed description of the teachers and classroom sizes).

Participating children were primarily White (86%) and spoke English as their primary language (97%). Based on parent report, participating children had the following difficulties: ADHD (4%), language impairment (3%), learning disability (3%) and behavioural problems (2%; Andrade & Tannock, 2013). Most participating parents were mothers (90%) and respondent parents had a range of education levels, including Grades 9-11 (2%), high school diploma or equivalent (6%), some college (10%), college graduate (35%), some university

(11%), university graduate (29%), and graduate degree (8%).

### 2.3.3 Measures

Parents and teachers completed the British version of the SDQ (R. Goodman, 1997). Parents were asked to think of the child's behaviour over the last 6 months and teachers were asked to think of the child's behaviour over the last 6 months or this school year. They were asked to answer all of the items, even if they were not absolutely certain. Ratings were scored as follows, based on instructions provided on the author's website ([www.sdqinfo.org](http://www.sdqinfo.org)): not true = 0; somewhat true = 1; certainly true = 2, with five of the items reverse-scored (not true = 2, certainly true = 0). Ratings within each subscale were summed to generate a subscale score and the total difficulties score was calculated by summing ratings on the four problem subscales (emotional symptoms, conduct problems, hyperactivity-inattention, and peer problems). Higher scores on these subscales indicate greater difficulties and positively-worded items within these subscales were reverse-scored. In contrast, higher scores on the prosocial behaviour scale indicate strengths.

### 2.3.4 Data Analysis

There were 465 children with complete teacher SDQ ratings, 22 with teacher ratings missing a single item, 1 with the teacher rating missing two items (from different subscales), and 13 with no teacher SDQ ratings available. There were 338 children with complete parent SDQ ratings, 8 children with parent SDQ ratings missing a single item, and 155 children with no parent SDQ ratings available.

Cronbach's alpha values for each of the five SDQ subscales and the total difficulties score were computed in SPSS 19 for parent and teacher ratings. Listwise deletion procedures were followed such that only participants with five valid item ratings within the subscale were

included.  Inter-rater Pearson correlations for the five subscales were examined using Mplus version 7 (L. K. Muthén & Muthén, 2012).  Maximum likelihood estimation with robust standard errors was used and the nesting of children within classrooms was accounted for.  Next, parent and teacher ratings were classified as normal, borderline, or abnormal for each subscale based on the scoring cut-offs for the British SDQ (www.sdqinfo.org).  Children falling within the borderline or abnormal range were considered "at risk" for difficulties in the area measured by the subscale, while children falling within the normal range were considered not at risk.  The percentages of children identified as at risk and not at risk by both raters as well as by one rater only were then cross-tabulated in SPSS 19.  Listwise deletion was used for the cross-tabulations of inter-rater agreement such that only participants with complete parent and teacher ratings were included in the analyses.

Separate CFAs were carried out for teacher and parent ratings in Mplus version 7 (L. K. Muthén & Muthén, 2012) using the robust weighted least squares estimator with a diagonal weight matrix (WLSMV), which has been recommended for use with ordered categorical data (Flora & Curran, 2004).  The WLSMV estimator uses pairwise deletion to handle missing data (L. K. Muthén & Muthén, 2012).  Children with no teacher ratings were dropped from the teacher analysis, resulting in an available sample size of 488.  Children with no parent ratings were dropped from the parent analysis, resulting in an available sample size of 346.  All item ratings were designated as categorical variables.  The nested design of the sample was accounted for in both parent and teacher analyses.  Composite reliability values were computed using the formula provided by Raykov (1997).

Model fit was evaluated based on the comparative fit index (CFI), the Tucker-Lewis Index (TLI) and the root mean square error of approximation (RMSEA; Schreiber, Stage, King,

Nora, & Barlow, 2006). For an acceptable fit, CFI and TLI > .90 and RMSEA < .08 were

required. For a good fit, CFI and TLI > .95 and RMSEA < .06 were required (Brown, 2006).

Factor loadings were interpreted according to criteria outlined by Comrey and Lee (1992), as

recommended by Tabachnick and Fidell (2007): a loading of .32 or more was required for the

variable to be interpreted; loadings > .71 were considered excellent, > .63 were considered very

good, > .55 were considered good, > .45 were considered fair, and > .32 were considered poor.

*P*-values (based on the standardized factor loading estimate divided by its standard error) were

examined and values < .05 were considered significant.

## 2.4 Results

Because a large number of parents did not provide SDQ ratings, children whose parents

did not return the SDQ were compared to those whose parents did. Children with and without

parent ratings did not differ in terms of teacher ratings on the SDQ emotional symptoms (Mann–

Whitney $U = 24,775.5$; p = .77), conduct problems (Mann–Whitney $U = 23,371.5$; p = .13),

hyperactivity–inattention (Mann–Whitney $U = 24,594.5$; p = .69), peer problems (Mann–

Whitney $U = 25,014.0$; p = .91), or prosocial behaviour (Mann–Whitney $U = 23,484.0$; p = .23)

subscales. However, children whose parents returned the SDQ had parents with significantly

more education (Mann–Whitney $U = 10,401.5$; p = .001) and were significantly more likely to be

in French immersion, $\chi^2 = 26.21(1)$, p < .001, than children whose parents did not return the

SDQ.

### 2.4.1 Internal Consistency

Cronbach's alpha values for teacher-rated subscales ranged from .65 (peer problems) to

.90 (hyperactivity-inattention) while for parent ratings the range was from .61 (conduct

problems) to .84 (hyperactivity-inattention). Composite reliability coefficients were above .70

for all teacher- and parent-rated subscales (see Table 2.1).  For teacher ratings, the lowest

composite reliability value was for the peer problems subscale (.79) and the highest was for

hyperactivity-inattention (.95).  For parent ratings, the lowest value was for emotional symptoms

(.80) and the highest was for hyperactivity-inattention (.92).  Because composite reliability is

calculated based on factor loadings, it was not possible to determine its value for the total

difficulties score.

**Table 2.1**

*Internal Consistency of SDQ Subscale and Total Difficulties Scores*

| Indicator | Emotional symptoms | Conduct problems | Hyperactivity -inattention | Peer problems | Prosocial behaviour | Total difficulties |
|---|---|---|---|---|---|---|
| Cronbach's alpha | | | | | | |
| Teacher | .79[a] | .75[b] | .90[c] | .65[d] | .83[e] | .88[f] |
| Parent | .67[g] | .61[h] | .84[i] | .63[g] | .72[g] | .82[j] |
| Composite reliability | | | | | | |
| Teacher | .91 | .92 | .95 | .79 | .91 | |
| Parent | .80 | .83 | .92 | .82 | .86 | |

[a] $n = 482$; [b] $n = 488$; [c] $n = 485$; [d] $n = 484$; [e] $n = 477$; [f] $n = 476$; [g] $n = 345$; [h] $n = 343$; [i] $n = 344$; [j] $n = 339$.

***2.4.2 Inter-rater Agreement***

Table 2.2 presents Pearson correlations between parent and teacher subscale scores.

Correlations ranged from .30 to .54 with the highest correlation found for the hyperactivity-

inattention subscale and the lowest for the prosocial behaviour subscale.  Next, the proportion of

children whose scores fell in the borderline or abnormal range was examined.  Further inter-rater

agreement analyses were carried out based on a subsample of participants who had both parent

and teacher ratings available ($n = 340$).  These participants did not differ from those with missing

data in terms of sex, $\chi^2 (1) = .21$, $p = .65$, or grade, $\chi^2 (2) = 3.67$, $p = .16$.  However, participants

with complete data had significantly more educated parents than participants with missing data

($n = 419$), Mann-Whitney $U = 10761.00$, $p = .001$.

Based on teacher ratings, 4.3%, 7.2%, 14.8%, 4.7%, and 7.4% of children fell within the

abnormal range for emotional symptoms, conduct problems, hyperactivity-inattention, peer

problems, and prosocial behaviour, respectively (3.3%, 4.5%, 3.5%, 3.7%, and 8.4% fell within

the borderline range, respectively).  According to parent ratings, 9.8%, 7.2%, 9.8%, 8.7%, and

2.0% of children fell within the abnormal range (7.2%, 9.8%, 6.1%, 6.9%, and 3.2% fell within

the borderline range, respectively).  When students were classified as at risk and not at risk

separately based on parent and teacher ratings, agreement across raters in terms of risk status was

found for over 80% of children for all of the subscales.  A portion of the sample, however,

ranging from 14% to 19% depending on the subscale, only met criteria for risk on the basis of

one rater (see Table 2.2).

**Table 2.2**

*Parent-Teacher Correlations and Agreement Rates for Risk Status*

| Subscale | Subscale score $r$[a] | Concordant risk status[b] | | Discordant risk status[b] | |
|---|---|---|---|---|---|
| | | Not at risk | At risk | Parent only | Teacher only |
| Emotional symptoms | .37 | 269 (79%) | 10 (3%) | 49 (14%) | 12 (4%) |
| Conduct problems | .46 | 262 (77%) | 16 (5%) | 43 (13%) | 19 (6%) |
| Hyperactivity-inattention | .54 | 253 (74%) | 30 (9%) | 25 (7%) | 32 (9%) |
| Peer problems | .31 | 273 (80%) | 15 (4%) | 37 (11%) | 15 (4%) |
| Prosocial behaviour | .30 | 285 (84%) | 7 (2%) | 11 (3%) | 37 (11%) |

*Note*. Due to rounding, percentages in rows may not sum to 100.

[a] $n = 494$; [b] $n = 340$.

### 2.4.3 Confirmatory Factor Analysis

***Teacher Ratings.***  All fit statistics for teacher ratings met recommended criteria for a

well-fitting model, $\chi^2(265) = 530.71$, $p < .001$; CFI = .966; TLI = .961; RMSEA = .045 (see

Table 2.3 for factor correlations and Table 2.4 for standardized factor loadings). All factor

loadings were significant ($p < .001$) and twenty of the twenty-five factor loadings fell within the

excellent range, with the others ranging from poor to very good. The strongest factor loadings

were found within the hyperactivity-inattention subscale (average loading = .90). The weakest

factor loadings were found within the peer problems subscale (average loading = .64). Both of

the items with poor factor loadings were from the peer problems subscale (Item 6: "Rather

solitary, tends to play alone" and item 23: "Gets on better with adults than with other children").

     *Parent Ratings.* For parent ratings, the CFI and RMSEA met recommended criteria for a

well-fitting model (CFI = .952; RMSEA = .035) and the TLI indicated an adequate fit (TLI =

.945), $\chi^2(265) = 376.12$, $p < .001$. Factor correlations are presented in Table 2.3 and Table 2.4

contains the standardized factor loadings. All factor loadings were significant ($p < .001$) and fell

within the good to excellent range, with the exception of item 3 in the emotional symptoms

subscale ("Often complains of headaches, stomach-aches or sickness") and item 6 in the peer

problems subscale ("Rather solitary, tends to play alone"), both of which had loadings falling

within the poor but still interpretable range. The strongest loadings were found for items making

up the hyperactivity-inattention subscale (average loading = .83), whereas the weakest were

found within the emotional symptoms subscale (average loading = .66). Factor loadings were

generally stronger for teacher ratings than for parent ratings, with the exception of the peer

problems subscale, for which parent ratings had a slightly stronger average factor loading.

**Table 2.3**

*Factor Correlations for Teacher and Parent SDQ Ratings*

| Subscale | Emotional symptoms | Conduct problems | Hyperactivity-inattention | Peer problems | Prosocial behaviour |
|---|---|---|---|---|---|
| Emotional symptoms | | .40*** | .37*** | .59*** | -.28*** |
| Conduct problems | .30*** | | .81*** | .76*** | -.78*** |
| Hyperactivity-inattention | .27** | .67*** | | .64*** | -.63*** |
| Peer problems | .60*** | .48*** | .44*** | | -.64*** |
| Prosocial behaviour | .01 | -.58*** | -.29*** | -.30*** | |

*Note:* Correlations for teacher ratings (*n* = 488) are presented above the diagonal. Correlations for parent ratings

(*n* = 346) are presented below the diagonal.

**p < .01; ***p < .001.

**Table 2.4**

*Standardized Factor Loadings and Standard Errors*

| Subscale and item | Teacher ratings | | Parent ratings | |
|---|---|---|---|---|
| | Standardized factor loading | SE | Standardized factor loading | SE |
| Emotional symptoms | | | | |
| Headaches, stomach-aches (3) | .62 | .07 | .37 | .09 |
| Worries (8) | .87 | .03 | .79 | .05 |
| Unhappy, downhearted (13) | .92 | .04 | .78 | .09 |
| Nervous in new situations (16) | .76 | .04 | .63 | .06 |
| Many fears (24) | .91 | .04 | .71 | .06 |
| Conduct problems | | | | |
| Temper tantrums (5) | .85 | .05 | .56 | .05 |
| [a]Obedient (7) | .82 | .04 | .82 | .06 |
| Fights or bullies (12) | .92 | .03 | .62 | .09 |
| Lies or cheats (18) | .84 | .05 | .71 | .05 |
| Steals (22) | .78 | .06 | .82 | .11 |
| Hyperactivity-inattention | | | | |
| Restless, overactive (2) | .94 | .01 | .87 | .03 |
| Fidgeting or squirming (10) | .94 | .02 | .91 | .02 |
| Easily distracted (15) | .86 | .02 | .86 | .03 |
| [a]Thinks before acting (21) | .88 | .02 | .68 | .05 |
| [a]Sees tasks through (25) | .88 | .02 | .84 | .02 |
| Peer problems | | | | |
| Plays alone (6) | .35 | .07 | .42 | .08 |
| [a]At least one friend (11) | .80 | .04 | .81 | .05 |
| [a]Generally liked (14) | .96 | .03 | .84 | .06 |
| Picked on or bullied (19) | .64 | .08 | .69 | .06 |
| Gets on better with adults (23) | .43 | .09 | .64 | .07 |
| Prosocial behaviour | | | | |
| Considerate of feelings (1) | 1.00 | .02 | .87 | .04 |
| Shares with others (4) | .79 | .04 | .75 | .05 |
| Helpful if someone is hurt (9) | .82 | .03 | .75 | .05 |
| Kind (17) | .78 | .04 | .69 | .06 |
| Volunteers to help (20) | .70 | .05 | .61 | .06 |

*Note.* SE = standard error.

[a] Indicates an item that is reverse-scored.

*p* < .001 for all parent and teacher factor loading

**2.5 Discussion**

This study sought to confirm the five-factor structure of parent and teacher ratings on the British SDQ using CFA in a sample of young Canadian children while also providing information about the measure's internal consistency and inter-rater reliability. Results suggest the hypothesized five-factor model fit the data for both parent and teacher ratings. In addition, the SDQ's five subscales demonstrated good internal consistency based on composite reliability coefficients and acceptable inter-rater agreement.

In separate CFAs, the five-factor model represented the data well for parent and teacher ratings. As in previous exploratory factor analysis and CFA studies (A. Goodman et al., 2010; R. Goodman, 2001; McCrory & Layte, 2012), the hyperactivity-inattention subscale had the strongest factor loadings for parents and teachers and the peer problems subscale had the weakest loadings for teacher ratings (A. Goodman et al., 2010; R. Goodman, 2001). Although the peer problems subscale has been found to exhibit the weakest factor loadings in prior CFAs of parent ratings (A. Goodman et al., 2010; McCrory & Layte, 2012), the present study found the emotional symptoms subscale to have the lowest loadings for parents. Within the peer problems subscale, the item "Rather solitary, tends to play alone" loaded poorly (although still significantly) on its factor for both parent and teacher ratings. In a previous CFA carried out in a sample of 9-year-old children, this item also showed a poor factor loading when rated by parents (McCrory & Layte, 2012). Collectively, these differences suggest that some factors may be more unitary than others and that a small number of items may load somewhat weakly on their subscales; however, the overall pattern of fit and the significant factor loadings in the present sample support the validity of the SDQ's five subscales as rated by parents and teachers of young children.

Factor correlations were generally stronger for teacher ratings than for parent ratings. None of the factor correlations was so strong as to suggest that the factors were not distinct (Kline, 2013); however, there were relatively strong correlations between conduct problems and hyperactivity-inattention in the current study for both parent and teacher ratings. This finding is consistent with the high level of comorbidity between ADHD and oppositional defiant disorder or conduct disorder (APA, 2000). Consistent with A. Goodman and colleagues (2010), the present study found evidence of a weak correlation between emotional symptoms and prosocial behaviour as rated by parents and teachers. This may indicate that, despite being prosocial, some children exhibit worries and unhappiness.

When internal consistency was measured with Cronbach's alpha, acceptable values were found for four of the teacher-rated subscales and the total difficulties score; however, the value for the peer problems subscale fell short of acceptable. Alpha values were lower for parent ratings, with only the hyperactivity-inattention, prosocial behaviour, and total difficulties scores having acceptable values. Similar alpha values have been reported in previous studies and the hyperactivity-inattention subscale has had the highest values for both parent and teacher ratings. Although prior studies have reported the peer problems subscale to have the lowest alpha value for both raters (A. Goodman et al., 2010; R. Goodman, 2001; McCrory & Layte, 2012), in the present sample the conduct problems subscale had the lowest value for parent ratings. Cronbach's alpha values in this study suggest that the internal consistency of the British SDQ in a sample of early elementary school Canadian children is comparable to that reported in previous studies.

While Cronbach's alpha values fell short of acceptable for one teacher-rated subscale and three parent-rated subscales, composite reliability values suggested good internal consistency for

all teacher- and parent- rated subscales.  Given the small number of items in the SDQ subscales, composite reliability may be a more accurate indicator of internal consistency than Cronbach's alpha (Graham, 2006).  As was the case with values for Cronbach's alpha, composite reliability values were generally slightly higher for teacher ratings than for parent ratings.  Based on a review of the literature, the present study is the first to use composite reliability as an indicator of internal consistency for the British version of the SDQ.

Inter-rater agreement for parents and teachers on all subscales was above the meta-analytic mean across measures of $r = .27$ (Achenbach et al., 1987) and inter-rater correlations in the present sample were stronger than those reported previously (A. Goodman et al., 2010; R. Goodman, 2001).  As in a previous study using the British SDQ, the weakest inter-rater correlation was for the prosocial behaviour subscale and the strongest was for hyperactivity-inattention (R. Goodman, 2001).  These results suggest that parents and teachers are more consistent in their ratings of hyperactivity-inattention and less consistent in their ratings of prosocial behaviour in a sample of young Canadian children.  It is not clear, however, whether differences in parent-teacher agreement are due to actual differences in behaviour across settings or to other factors such as raters' perceptions (De Los Reyes et al., 2009).  Overall, the SDQ performed adequately in terms of inter-rater reliability.

The percentage of children who fell within the borderline and abnormal ranges for parent and teacher ratings in the current sample was somewhat lower than the anticipated 10% for each category (R. Goodman, 1997).  Teachers rated more children as within the abnormal range for hyperactivity-inattention than did parents but otherwise parents rated more children as falling within the borderline or abnormal ranges for the remaining problem subscales.  Similar findings in terms of rates of elevated scores and the number of children identified by parent- versus

teacher-ratings were reported in a sample of Canadian children based on a diagnostic interview (Breton et al., 1999). The concordance of teacher and parent ratings when children were classified as at risk for difficulties in any of the five domains sampled by the SDQ if they fell within the borderline or abnormal range of scores was also examined. Although children in the borderline range have been grouped with children in the normal range in previous studies using the British SDQ (R. Goodman, 2001), other studies have grouped those within the borderline and abnormal ranges together (Conti-Ramsden, Mok, Pickles, & Durkin, 2013). The latter approach was used here as there is evidence that increases in symptom ratings on the SDQ are associated with increases in impairment, whether or not symptom levels fall within the abnormal range (Huculak & McLennan, 2014). Moreover, a key purpose of screening is to identify difficulties as early as possible and begin delivering intervention (Dowdy, Ritchey, et al., 2010); thus, identifying only children who are already within the clinical range would miss those children with subthreshold levels of difficulty. Results suggest that parents and teachers largely agree about children's risk status, particularly when children are *not* displaying problematic levels of behaviour. However, many children were also identified as at risk by one rater only (more commonly parents than teachers), supporting previous recommendations that a multi-informant approach should be used when assessing children (Johnston & Murray, 2003).

### 2.5.1 Implications for Practice

In a sample of young Canadian early elementary school children, parent and teacher ratings on the British version of the SDQ perform adequately in terms of factor structure, internal consistency, and inter-rater reliability. Thus, Canadian practitioners may wish to include the SDQ in assessments of early elementary school children as a means of gathering information about their social, emotional, behavioural, and attentional functioning. School districts outside

of Canada have collected SDQ ratings for all children (Lane et al., 2011; White et al., 2013) and other schools or districts may wish to collect similar information as part of a universal screening battery. This information may enhance communication between school professionals about children's social and emotional functioning (White et al., 2013) or may be useful in identifying children who would benefit from targeted supports within or outside of the school setting (Lane et al., 2011). There has been interest in the literature in identifying the best source of information on children's functioning (Dirks et al., 2011). While the predictive value of parent or teacher ratings in the current sample was not determined, results suggest that parents and teachers may view young children's difficulties differently. Therefore, it is prudent to gather information from both raters when possible to avoid missing children who may be at risk for mental health-related difficulties.

### 2.5.2 Limitations

The results of the present study must be interpreted in light of its limitations. First, as the participants were members of a convenience sample, there is a need to replicate the results with a larger and more representative sample of Canadian children, including participants who are more diverse in terms of language, cultural background, and socioeconomic status. In addition, the age range was restricted to primary grade children (ages 6 to 9) and the SDQ is intended for use with children ages 4 to 16. Therefore, the results may not generalize to younger and older children or adolescents. Additionally, parent ratings were missing for many participants. Although there were no significant differences between participants with and without parent data on the teacher SDQ subscale ratings, those participants with parent SDQ ratings had parents with higher levels of education and were more likely to be in French immersion than participants whose parents did not complete the SDQ. These differences may have affected the

representativeness of the sample used in the parent analysis.  In addition, the SDQ has not been

normed on a Canadian sample, making it necessary to rely on normative information from the

United Kingdom or the United States.  Information derived from the SDQ for Canadian children

should be interpreted with this in mind.  Finally, Canada is a diverse country with many

languages represented, including French, the country's second official language.  The present

study used only the British version of the SDQ; thus, results cannot be generalized to translations

of the SDQ that may be used with speakers of other languages within the Canadian context.

### 2.5.3 Conclusion

Early intervention is critical in the domain of children's mental health in order to prevent

long-term negative outcomes and universal screening has been suggested as one means of

facilitating early access to services for young children (Dowdy et al., 2013).  Such universal

screening efforts require measures that are reliable, valid, and appropriate for the context (e.g.,

that do not overburden the respondent).  This preliminary investigation supports the reliability of

the SDQ as a measure of Canadian early elementary school children's social, emotional,

behavioural, and attentional functioning based on parent and teacher ratings.  In the current

sample, Cronbach's alpha values were much lower than composite reliability values, another

indicator of internal consistency that may be more appropriate for use with the SDQ given the

small number of items in each subscale.  Computing the composite reliability of the subscales in

future studies may provide a more accurate estimate of the SDQ's internal consistency.  The

five-factor structure was replicated in the present sample for parent and teacher SDQ ratings.

Future studies should examine the factor structure for Canadian children across a broader range

of ages and using a larger and more heterogeneous sample.  In addition, future research could

assess the utility of the SDQ for universal screening to identify children in need of a more

comprehensive assessment (Lane et al., 2011).  It would also be useful to carry out similar analyses of self-report ratings for older children and adolescents, as universal screening for mental health issues has been recommended at the high school level (Kalberg, Lane, Driscoll, & Wehby, 2011).  Collecting parent and teacher ratings on the SDQ may help school districts to better understand the mental health needs of their students and to assess the impact of school-wide efforts to reduce mental health difficulties (Lane, Menzies, Ennis, & Bezdek, 2013).

**CHAPTER THREE**

**Incremental Validity of Teacher and Parent Symptom and Impairment Ratings when**

**Screening for Mental Health Difficulties**

**3.1 Abstract**

Although universal screening for mental health difficulties is increasingly recognized as a way to

identify children who are at risk and provide early intervention, little research exists to inform

decisions about screening, such as the choice of informants and the type of information collected.

The present study examined the incremental validity of teacher- and parent-rated (primarily

mothers) symptoms and impairment in a non-referred sample of early elementary school children

($n$ = 320, ages 6 to 9) in terms of predicting teacher-rated impairment one year later. Teacher-

rated symptoms and impairment and parent-rated impairment were each unique predictors of

later impairment; however, parent-rated symptoms did not contribute to the prediction of later

impairment above and beyond these other indicators. The results indicate that, when screening

for mental health difficulties in the school system, impairment ratings collected across settings

add useful information but it may not be necessary to use parent symptom ratings when teacher

symptom ratings are available. In addition, a supplemental analysis highlights the importance of

attending to criterion contamination in studies of incremental validity.

**3.2 Introduction**

Identifying mental health difficulties early and providing appropriate intervention may be more effective than later treatment and may prevent more serious problems from developing (Dodge & Pettit, 2003; Durlak & Wells, 1998).  For these reasons, researchers, clinicians, and policymakers have recently made early intervention a priority in the area of children's mental health services (Bringewatt & Gershoff, 2010; Herrman, 2014; Mental Health Commission of Canada, 2012; World Health Organization, 2013).  Results of early intervention studies are encouraging (Dadds et al., 1999; Dodge et al., 2015).  For example, large effect sizes have been reported for early intervention programs targeting sub-clinical externalizing problems, whereas clinical-level externalizing is less amenable to intervention (Durlak & Wells, 1998).  The identification of elevated but sub-clinical difficulties is essential for early intervention and prevention (Durlak & Wells, 1998) and, therefore, proactive approaches are necessary to ascertain which children may be at risk.

Schools are uniquely positioned to play an important role in identifying children with mental health needs (Bringewatt & Gershoff, 2010; Durlak & Wells, 1998).  Although universal screening has been recommended as a means of identifying mental health difficulties early within the school system (Dever et al., 2015; Dowdy, Ritchey, et al., 2010; Glover & Albers, 2007), schools have generally taken a reactive approach to mental health (Albers et al., 2007; Lane et al., 2012).  Encouragingly, there has been a recent increase in mental health screening in school settings (Lane et al., 2012) and universal screening for mental health difficulties is now recognized as "current best practice" (Dever et al., 2015, p. 10).  However, relatively little research exists to guide the implementation of screening practices within the school system (Dever et al., 2015; Dowdy & Kim, 2012).

While there are a number of measures with demonstrated reliability and validity for mental health screening (for reviews, see Feeney-Kettler et al., 2010; Jenkins et al., 2014; Levitt et al., 2007), few studies have examined the value of gathering information from multiple informants or the relative usefulness of different types of information that can be collected, such as symptom severity or impairment ratings. It should not be assumed that more is better in terms of assessment as brief, simple approaches could lead to the same conclusions as more comprehensive methods; thus, clinicians and researchers should provide evidence of the incremental validity of assessment practices (Johnston & Murray, 2003). Nowhere is this more practically important than in the context of school-based screening, where efficiency is essential due to the large scale of universal screening and the limited resources within the school system (Glover & Albers, 2007). The present study focuses on the incremental validity of teacher and parent ratings of symptoms and impairment as part of a brief screen for common childhood mental health difficulties in young children, including internalizing (i.e., anxiety and depression), externalizing (i.e., conduct problems and oppositional defiance) and inattention/hyperactivity. Given that self-reports of symptoms demonstrate limited reliability in young children (Edelbrock, Costello, Dulcan, Kalas, & Conover, 1985), child self-report is not considered here.

### 3.2.1 Incremental Validity of Multiple Informants

There is currently little research to guide the choice of informants for mental health screening in school settings, leaving practitioners to rely on "clinical intuition and judgment" (Dowdy & Kim, 2012, p. 1). Practical considerations, such as the cost associated with gathering ratings from various informants, may also influence the selection of informants (Glover & Albers, 2007). As a result, teacher ratings are often used for school-based screening (Lane et al., 2011; White, Connelly, Thompson, & Wilson, 2013), in part because they are less costly than

parent ratings (Flanagan, Bierman, & Kam, 2003).  Several studies have demonstrated the validity of teacher ratings for a variety of mental health difficulties in elementary school children (Epkins, 1993; Hill et al., 2004; Sharp et al., 2005; Verhulst et al., 1994).  Therefore, the use of teacher ratings for mental health screening is both practical and valid.

Despite the validity of teacher ratings, agreement between parent and teacher ratings is generally low (Achenbach et al., 1987; De Los Reyes et al., 2015; Offord et al., 1996).  For example, a recent meta-analysis found mean correlations between parent and teacher ratings of $r$ = .21 for internalizing symptoms and $r$ = .28 for externalizing symptoms (De Los Reyes et al., 2015).  There is also evidence that discrepancies between parent and teacher ratings are due to actual differences in the child's behaviour across settings (De Los Reyes, Henry, Tolan, & Wakschlag, 2009; Dirks, De Los Reyes, Briggs-Gowan, Cella, & Wakschlag, 2012).  In addition, parent and teacher ratings may vary in their sensitivity to different symptoms, with teacher ratings of externalizing being more informative than parent ratings and parent ratings of internalizing more informative than teacher ratings (R. Goodman et al., 2000).  Parents' screening ratings also identify more children as at-risk in most areas than do teachers' ratings (Aitken, Martinussen, Wolfe, & Tannock, 2015).  Given these differences, it may be necessary to collect ratings from both parents and teachers in order to identify all children who are at risk for mental health difficulties.

It is widely agreed that multiple informants are necessary for comprehensive mental health assessments of children (Hunsley & Mash, 2007; Sowerby & Tripp, 2009).  Several studies have reported that using both parent and teacher ratings predicts a concurrent psychiatric diagnosis more accurately than using a single informant (R. Goodman et al., 2000; Johnson et al., 2014; Power et al., 1998).  Similar findings have been reported in the few studies that have

examined parent and teacher ratings as predictors of outcomes other than diagnosis. For example, parent-rated externalizing in kindergarten predicts internalizing and externalizing in Grade 1, above and beyond teacher-rated externalizing in non-referred children (Lochman & The Conduct Problems Prevention Research Group, 1995). Similarly, Hill and colleagues (2004) reported that using both parent and teacher externalizing ratings in Grade 1 resulted in the best prediction of Grade 4 and 5 outcomes. Finally, in a study that screened for a broader set of mental health difficulties in children ages 4 to 16, both parent and teacher ratings predicted a range of negative outcomes six years later (Verhulst et al., 1994). Therefore, collecting both parent and teacher ratings may be useful when screening for mental health difficulties.

While previous studies generally support the value of multi-informant ratings, it remains unclear for several reasons whether multiple informants should be used in mental health screening in school settings. First, previous studies have one or more methodological limitations that make it difficult to draw firm conclusions about the incremental validity of multiple informants (De Los Reyes et al., 2015). In particular, the majority of these studies have used criterion variables that are based on one or more of the predictor variables, which may overestimate the association between the predictor and the criterion (referred to as "criterion contamination"; De Los Reyes et al., 2015). Second, much of the research on multiple informants has focused on psychiatric diagnosis as an outcome; however, these findings may not generalize to screening given that screening typically seeks to identify sub-clinical levels of difficulty (i.e., at-risk status). Moreover, unlike comprehensive psychological assessment, screening requires efficient practices (Dowdy & Kim, 2012; Glover & Albers, 2007) and is generally intended to identify children in need of more intensive services, including in-depth assessment (R. Goodman et al., 2000). Finally, in the few studies that have examined sub-

clinical outcomes (Hill et al., 2004; Lochman & The Conduct Problems Prevention Research Group, 1995; Verhulst et al., 1994), the use of lengthy questionnaires or the focus on externalizing behaviour alone make it difficult to generalize the results to universal screening for mental health difficulties.  Therefore, although it has been suggested that multiple informants may not be necessary for screening purposes (Dowdy & Kim, 2012), further research is necessary to test this hypothesis.

### 3.2.2 Incremental Validity of Symptom and Impairment Ratings

Current approaches to screening generally focus on symptom ratings (R. Goodman, 1999); however, symptoms and impairment are related but distinct constructs and may therefore contribute unique information when screening for mental health difficulties (Gadow et al., 2013; Rapee et al., 2012).  In particular, including impairment ratings may result in more accurate assessment and screening.  Several studies have demonstrated unrealistically high rates of disorders when symptoms alone are considered and these rates are significantly reduced when impairment is required (for a review, see Rapee et al., 2012).  For example, in a recent study of children ages 5 to 8 in the Netherlands, 31% met diagnostic criteria based on symptoms only; however, this rate fell to 23% when mild impairment was required, which is more consistent with prevalence estimates (Rijlaarsdam et al., 2015).  Failing to consider impairment is particularly likely to lead to overestimates of disorder rates in community samples (Rapee et al., 2012); therefore, including impairment ratings when screening for mental health difficulties may reduce false positives.

Impairment ratings may also help with the issue of under-identification of children at risk given that several studies have identified groups of children who are impaired without exhibiting clinical levels of symptoms.  In non-referred samples, between 6% and 13% of children and

adolescents are significantly impaired but do not meet diagnostic criteria based on symptoms (Angold, Costello, Farmer, et al., 1999; Wille et al., 2008).  Similar findings have been reported in clinic-referred samples (Gadow et al., 2013; Huculak & McLennan, 2014).  For example, in a sample of children and youth referred for school-based mental health services, 13% were significantly impaired but would not have been considered at-risk if only symptoms were assessed (Huculak & McLennan, 2014).  These findings suggest that collecting impairment ratings in addition to symptom ratings may be especially useful in a screening context given the emphasis on sub-clinical difficulties and on identifying as many children as possible who are at risk (R. Goodman, 2001).

Finally, impairment ratings demonstrate prognostic value and predict a range of outcomes over time.  For example, Costello, Angold, and Keeler (1999) found that children who met symptom criteria (but not full diagnostic criteria) were at an increased risk for serious emotional disturbance in adolescence only if they were impaired in childhood.  Other studies have examined baseline levels of impairment as a predictor of future outcomes using regression analyses.  In one such study, parent and teacher ratings of impairment at baseline (considered in separate analyses) predicted both mental health service use and self-harm three years later in a community sample of children and youth (Stringaris & Goodman, 2013).  Similarly, Pickles and colleagues (2001) found that initial impairment predicted externalizing symptoms and diagnoses 19 months later in a community sample of children and adolescents, above and beyond baseline symptom levels.  However, impairment was not predictive of future depression symptoms or diagnosis (Pickles et al., 2001).  Nevertheless, impairment appears to contribute important information that may help to predict the persistence of some difficulties over time.  Given the varying demands across home and school contexts, examining the prognostic value of parent and

teacher impairment ratings may be informative for mental health screening.

### 3.2.3 Current Study

The present study seeks to determine the incremental validity of gathering parent and teacher impairment ratings and parent symptom ratings in addition to the usually-collected teacher symptom ratings when screening for mental health difficulties in young children. It is hypothesized that teacher and parent ratings of symptoms and impairment will each contribute uniquely to the prediction of children's functioning one year later. Because boys may be more likely than girls to be rated as impaired and to be false positives in mental health screening (R. Goodman et al., 2000; Wille et al., 2008), sex should also be controlled for in the analyses.

The criterion variables used in studies of incremental validity can have a significant influence on their results (De Los Reyes et al., 2015; Hunsley, 2003; Johnston & Murray, 2003). It is important that the criterion variable is clinically relevant, reliable, and that it is not based on information from any of the predictor variables in order to avoid criterion contamination (De Los Reyes et al., 2015; Hunsley, 2003). As a result, the present study used ratings of children's impairment one year later (provided by a different teacher) as the primary criterion. Because the focus is on school-based mental health and screening, children's functioning at school is a clinically useful indicator.

## 3.3 Method

### 3.3.1 Procedure

This study uses data from a two-year prospective study of attention and academic outcomes in a non-referred, school-based sample of children (for additional details, see Andrade & Tannock, 2013; Normand, Flora, Toplak, & Tannock, 2012). The study was approved by research ethics boards at the University of Toronto and the Hospital for Sick Children. Children

from one rural and one suburban Ontario school board were included in the sample if they were in a mainstream classroom in English or French immersion, had no sensory or physical impairments that would affect their ability to complete the other study tasks or hear instructions, had written parental and teacher informed consent and gave verbal assent (Normand et al., 2012). For the purposes of the present study, predictor variables were measured in November (Time 1) and criterion variable measures were collected approximately one year later (Time 2).

### 3.3.2 Sample

Seven schools took part in the study and all Grade 1 to 3 teachers ($N = 51$) in these schools consented to participate (Normand et al., 2012 provides details about the teachers and classroom sizes). The parents of 501 children ranging in age from 6 to 9 years ($M = 7.5$ years) consented to participate. Approximately one quarter of the children were in a French immersion program in which French was the main language of instruction. The majority were White (86%) and spoke English as their primary language (97%). Most respondent parents were mothers (90%) and parental education was slightly higher in the present sample than across the Canadian population, with 72% of respondent parents holding a college or university degree.

### 3.3.3 Measures

**Symptoms.** Parents and teachers completed the SDQ (R. Goodman, 1997), a 25-item behavioural screening measure. Each item is rated on a Likert-type scale (not true, somewhat true, certainly true) and the items are divided equally into five subscales: emotional symptoms, conduct problems, hyperactivity-inattention, peer problems, and prosocial behaviour. The first four subscales can be combined into a total difficulties score, with higher scores indicating more symptoms. In the present study, Time 1 parent and teacher total difficulties scores were used as indicators of symptom severity. The British version of the SDQ was used regardless of whether

the child was in French immersion.  The SDQ is freely available at www.sdqinfo.org.  Its reliability and validity have been demonstrated in community and clinic samples in multiple countries (Stone et al., 2010; Woerner et al., 2004) and the adequacy of its factor structure, internal consistency and inter-rater agreement for parents and teachers have also been demonstrated in a sample of Canadian children (Aitken, Martinussen, Wolfe, et al., 2015).

*Impairment.*  The SDQ has an optional impact supplement consisting of a second page of questions about the impairment the child experiences in a variety of contexts (R. Goodman, 1999).  Informants are asked whether the child experiences difficulties with emotions, concentration, behaviour, or getting along with others and, if so, the degree to which the difficulties distress the child and interfere with home life, friendships, classroom learning, and leisure activities (teachers rate only the child's distress and interference with peer relationships and classroom learning).  Each item is rated on a four-point Likert-type scale (not at all, only a little, quite a lot, a great deal).  Ratings of not at all and only a little receive a score of 0, quite a lot is scored as 1, and a great deal is scored as 2.  These items are summed to generate an impact score ranging from 0 to 10 for parents and 0 to 6 for teachers.  In the present study, parent and teacher impact scores at Time 1 were used as indicators of impairment.  The SDQ impact score demonstrates high internal consistency for both parent and teacher ratings (Stringaris & Goodman, 2013).

*Criterion Variables.*  The primary criterion variable was Time 2 teacher-rated impairment based on the SDQ impact supplement, referred to here as Time 2 impairment to distinguish it from the predictor variables.  Because children had advanced a grade, different teachers provided ratings at Time 1 and Time 2 and, therefore, the predictor variables and Time 2 impairment are not based on shared information.  A small number of participants (4%) had the

same teacher at both time points and, to ensure that this was not influencing the results, the analyses were run with and without these participants.

As a supplemental analysis, risk for a mental health diagnosis at Time 2 was used as a criterion variable. Risk for diagnosis was determined using the SDQ computer scoring algorithm (www.sdqinfo.org), which combines parent and teacher ratings of emotional symptoms, conduct problems, hyperactivity-inattention, and impairment to determine the likelihood that an individual meets criteria for any common mental health diagnosis (anxiety-depressive disorder, conduct-oppositional disorder, or hyperactivity-inattention disorder; R. Goodman, Ford, Corbin, & Meltzer, 2004). When ratings from one informant are missing, the algorithm uses the ratings from the remaining informant. The algorithm generates values of unlikely, possible, and probable diagnosis and, in the present study, the possible and probable classifications were collapsed into a single at risk category, while participants for whom a diagnosis was unlikely were considered at low risk. The utility of the SDQ scoring algorithm in screening for psychiatric disorders has been demonstrated in a large community sample of children and adolescents (R. Goodman et al., 2000).

### 3.3.4 Data Analysis

The primary analysis was a negative binomial regression with Time 2 impairment as the dependent variable. Negative binomial regression was used due to the strong positive skew of the dependent variable, including a high frequency of zeroes that were not believed to be zero-inflated, and the variance being much larger than the mean, suggesting overdispersion (Allison, 2012; Gardner, Mulvey, & Shaw, 1995). Potential control (child age and sex, parent education) and predictor variables (Time 1 teacher- and parent-rated symptoms and impairment) were examined to determine whether they were associated with Time 2 impairment. Variables that

were significantly related to this criterion variable were included in the regression in a way that approximated gathering increasing amounts of screening data. Block 1 included Time 1 teacher-rated symptoms (along with control variables) and Block 2 added Time 1 teacher-rated impairment because these ratings would be the least onerous to collect if teachers were already rating symptoms. In Block 3, Time 1 parent-rated impairment was added (because of its relatively low demands in terms of scoring and parents' time), followed by Time 1 parent-rated symptoms in Block 4. Incidence rate ratios (IRRs) were calculated following Kremelberg (2011). The same variables were also considered as predictors of Time 2 risk for diagnosis (at risk vs. low risk) in a supplemental analysis using binary logistic regression. Alpha was set at .05 for all analyses.

**3.4 Results**

Data were missing on one or more variables for 181 participants, mainly due to missing parent-rated symptoms and/or impairment. This resulted in a sample of 320 participants (161 boys, 159 girls) with complete data. These participants did not differ from those with missing data in terms of age, sex, or either of the criterion variables (Time 2 impairment, Time 2 risk for diagnosis). The analyses reported below are based on the 320 participants with data available on parent education, Time 1 teacher- and parent-rated symptoms and impairment, and Time 2 impairment and risk for diagnosis. Time 2 risk for diagnosis was calculated based on both parent and teacher symptom and impairment ratings for 85% of the 320 participants; however, because the SDQ algorithm allows for data to be missing for one rater, risk for diagnosis was calculated based only on teacher symptom and impairment ratings for 14% of participants (the remainder were missing parent impairment ratings only). There were 26 participants missing age data and the mean age within their grade was substituted. Descriptive statistics for the independent and

dependent variables are presented in Table 3.1.

**Table 3.1**

*Descriptive Statistics and Spearman Correlations among Criterion and Predictor Variables*

| Variable | 2 | 3 | 4 | 5 | 6 | 7 | Range | Mean | SD |
|---|---|---|---|---|---|---|---|---|---|
| 1. Time 2 impairment[a] | .03 | -.18** | .46*** | .58*** | .36*** | .46*** | 0-6 | 0.49 | 1.15 |
| 2. Age | – | -.02 | -.02 | .05 | .01 | .07 | 5.95-8.96 | 7.54 | 0.87 |
| 3. Parent education[b] | | – | -.13* | -.10 | -.20*** | -.03 | 2-8 | 5.78 | 1.43 |
| 4. Time 1 teacher-rated symptoms[c] | | | – | .59*** | .41*** | .38*** | 0-30 | 5.61 | 5.70 |
| 5. Time 1 teacher-rated impairment[a] | | | | – | .38*** | .48*** | 0-5 | 0.34 | 0.88 |
| 6. Time 1 parent-rated symptoms[d] | | | | | – | .48*** | 0-29 | 7.36 | 5.17 |
| 7. Time 1 parent-rated impairment[a] | | | | | | – | 0-7 | 0.47 | 1.26 |

[a] Scores $\geq 1$ are considered Borderline and $\geq 2$ are considered Abnormal; [b] Coded on an 8-point scale ranging from 1 = Grades 0-8 to 8 = Post-college degree

(Master's); [c] Scores $\geq 12$ are considered Borderline and $\geq 16$ are considered Abnormal; [c] Scores $\geq 14$ are considered Borderline and $\geq 17$ are considered

Abnormal.

*p* < .05; **p* < .01; ***p* < .001.

### 3.4.1 Primary Analysis

Associations between Time 2 impairment and potential predictor variables were examined using Spearman correlations (see Table 3.1). Parent education, teacher-rated symptoms and impairment, and parent-rated symptoms and impairment were each significantly correlated with Time 2 impairment and were therefore included as predictors in the regression analysis. Because age was not significantly correlated with Time 2 impairment, age was not controlled for in the analysis. Based on a Mann-Whitney $U$ test, boys had significantly greater Time 2 impairment than girls ($p < .001$) and, therefore, sex was included as a control variable in the regression analysis.

In Block 1 of the regression, teacher-rated symptoms were entered along with children's sex and parent education. Male sex, $B = 1.06$, Wald $\chi^2 = 15.69$, $p < .001$, IRR = 2.88, 95% CI [1.71, 4.85] and teacher-rated symptoms, $B = 0.16$, Wald $\chi^2 = 72.28$, $p < .001$, IRR = 1.17, 95% CI [1.13, 1.22] were both significant predictors of Time 2 impairment but parent education was not, $B = -0.08$, Wald $\chi^2 = 0.88$, $p = .35$, IRR = 0.92, 95% CI [0.78, 1.09].

In Block 2, Time 1 teacher-rated impairment was added to the model and predicted Time 2 impairment, $B = 0.29$, Wald $\chi^2 = 4.44$, $p = .04$, IRR = 1.34, 95% CI [1.02, 1.75], above and beyond male sex, $B = 1.09$, Wald $\chi^2 = 16.09$, $p < .001$, IRR = 2.99, 95% CI [1.75, 5.10], parent education, $B = -0.07$, Wald $\chi^2 = 0.70$, $p = 0.40$, IRR = 0.93, 95% CI [0.79, 1.10], and teacher-rated symptoms, $B = 0.12$, Wald $\chi^2 = 20.18$, $p < .001$, IRR = 1.13, 95% CI [1.07, 1.19].

Time 1 parent-rated impairment was added to the model in Block 3 and predicted Time 2 impairment, $B = .22$, Wald $\chi^2 = 7.59$, $p = .006$, IRR = 1.25, 95% CI [1.07, 1.46], above and beyond male sex, $B = 1.16$, Wald $\chi^2 = 17.35$, $p < .001$, IRR = 3.19, 95% CI [1.85, 5.50], parent education, $B = -0.09$, Wald $\chi^2 = 1.09$, $p = .30$, IRR = 0.91, 95% CI [.77, 1.08], teacher-rated

symptoms, $B = 0.08$, Wald $\chi^2 = 8.23$, $p = .004$, IRR = 1.09, 95% CI [1.03, 1.15], and Time 1 teacher-rated impairment, $B = 0.30$, Wald $\chi^2 = 4.39$, $p = 0.04$, IRR = 1.35, 95% CI [1.02, 1.78].

In Block 4, parent-rated symptoms were added to the model (see Table 3.2). Parent-rated symptoms did not significantly predict Time 2 impairment above and beyond sex, parent education, Time 1 teacher-rated symptoms and impairment and Time 1 parent-rated impairment. However, Time 1 teacher-rated symptoms and impairment and Time 1 parent-rated impairment remained significant predictors of Time 2 impairment. The IRRs indicate that, holding other variables constant, boys experienced 3.18 times more impairment at Time 2 than girls, a one-point increase in teacher-rated symptoms was associated with an 8% increase in Time 2 impairment, and a one-point increase in Time 1 teacher-rated impairment was associated with a 38% increase in Time 2 impairment. A one-point increase in Time 1 parent-rated impairment was associated with a 21% increase in Time 2 impairment. The same pattern of results was found when the analysis was re-run excluding the 14 participants who had the same teacher at Time 1 and Time 2; therefore, these participants were retained in the analysis.

Next, four interaction terms were entered in Block 5 to examine moderation effects of sex on Time 1 teacher-rated symptoms and impairment and Time 1 parent-rated symptoms and impairment. None of the interactions reached or approached significance and, as a result, they were removed from the model.

**Table 3.2**

*Negative Binomial Regressions Predicting Time 2 Impairment*

| Predictor | $B$ (SE) | Wald $\chi^2$ | IRR |
|---|---|---|---|
| Full model | | | |
| Child sex (male) | 1.16 (0.28) | 17.37*** | 3.18 |
| Parent education | -0.08 (0.09) | 0.77 | 0.92 |
| Time 1 teacher-rated symptoms | 0.08 (0.03) | 5.41* | 1.08 |
| Time 1 teacher-rated impairment | 0.32 (0.15) | 4.72* | 1.38 |
| Time 1 parent-rated impairment | 0.19 (0.09) | 3.99* | 1.21 |
| Time 1 parent-rated symptoms | 0.02 (0.03) | 0.38 | 1.02 |
| Model 1A | | | |
| Child sex (male) | 1.34 (0.25) | 28.71*** | 3.80 |
| Parent education | -0.10 (0.08) | 1.42 | 0.91 |
| Time 1 parent-rated symptoms | 0.13 (0.02) | 42.78*** | 1.14 |
| Model 2A | | | |
| Child sex (male) | 1.31 (0.26) | 26.09*** | 3.72 |
| Parent education | -0.15 (0.08) | 3.26 | 0.86 |
| Time 1 parent-rated symptoms | 0.06 (0.03) | 5.31* | 1.06 |
| Time 1 parent-rated impairment | 0.36 (0.09) | 17.82*** | 1.44 |
| Model 3A | | | |
| Child sex (male) | 1.26 (0.27) | 21.51*** | 3.53 |
| Parent education | -0.07 (0.09) | 0.63 | 0.93 |
| Time 1 parent-rated symptoms | 0.05 (0.03) | 3.23 | 1.05 |
| Time 1 parent-rated impairment | 0.21 (0.09) | 4.97* | 1.23 |
| Time 1 teacher-rated impairment | 0.57 (0.11) | 26.95*** | 1.77 |

*Note.* IRR = incidence rate ratio.

*$p < .05$. **$p < .01$. ***$p < .001$.

Finally, an alternate order of entry was tried, with parent ratings entered before teacher ratings. First, parent-rated symptoms were entered along with sex and parent education (see Model 1A in Table 3.2). In this model, parent-rated symptoms were a significant predictor of Time 2 impairment. Next, Time 1 parent-rated impairment was added (see Model 2A in Table 3.2), resulting in both Time 1 parent-rated symptoms and impairment being significant predictors of Time 2 impairment. Finally, Time 1 teacher-rated impairment was added to the model (see Model 3A in Table 3.2), at which point parent-rated symptoms were no longer a significant predictor of Time 2 impairment. These findings indicate that parent-rated symptoms at Time 1 are not a significant unique predictor of Time 2 impairment, perhaps due to shared variance between parent-rated symptoms and impairment and between parent- and teacher-rated symptoms.

### 3.4.2 Supplemental Analysis

The overall likelihood of any diagnosis at Time 2 was used as the criterion variable in a supplemental binary logistic regression analysis (24% of participants were at risk for a diagnosis and 76% were low risk). Kruskal-Wallis tests indicated that age was not significantly associated with Time 2 risk for diagnosis ($p = 0.84$) but parent education ($p = 0.04$), Time 1 teacher-rated symptoms and impairment, and Time 1 parent-rated symptoms and impairment were (all $ps < .001$). There were also significant sex differences such that boys were more likely than girls to be at risk for a diagnosis, $\chi^2(1) = 13.91$, $p < .001$. Therefore, sex, parent education, Time 1 teacher-rated symptoms and impairment, and Time 1 parent-rated symptoms and impairment were included as predictors of Time 2 risk for diagnosis. When all six predictors were considered simultaneously, only Time 1 teacher-rated impairment, $B = 0.83$, $p = 0.02$, OR = 2.28, 95% CI [1.18, 4.54], Time 1 parent-rated impairment, $B = 0.51$, $p = 0.01$, OR = 1.66, 95%

CI [1.13, 2.44], and Time 1 parent-rated symptoms, $B = 0.15$, $p = .001$, OR = 1.17, 95% CI [1.07, 1.27] were significant predictors of Time 2 risk for diagnosis. Sex ($p = .10$) and parent education ($p = 0.49$) were not significant predictors and teacher-rated symptoms fell just above the cut-off for significance, $B = 0.09$, $p = .053$, OR = 1.10, 95% CI [0.999, 1.20]. It is important to note that in this analysis, parent ratings of symptoms and impairment contributed to both the predictor and criterion variables. Although teacher-rated symptoms and impairment were also part of the criterion variable, different teachers provided ratings at Time 1 and Time 2 for 96% of participants. This differs from the primary analysis in which neither parents nor the same teachers (except for 4% of participants) contributed to both the predictor and criterion variables.

## 3.5 Discussion

This study examined the incremental validity of teacher and parent ratings of symptoms and impairment in terms of predicting functioning one year later in a non-referred sample of children. The goal was to simulate school-based screening for mental health difficulties and determine the value of gathering information that goes beyond teacher ratings of symptoms, which are typically collected for school-based purposes. With this goal in mind, Time 2 teacher-rated impairment was selected as the primary criterion variable given its relevance to school-based intervention and the fact that different teachers provided predictor and criterion ratings. It was anticipated that teacher- and parent-rated symptoms and impairment at Time 1 would each contribute uniquely to the prediction of functioning one year later and the results partially support this hypothesis. In the primary analysis, teacher-rated symptoms, as well as teacher- and parent-rated impairment at Time 1, were all significant predictors of Time 2 impairment when considered in a model that also included parent-rated symptoms. However, parent-rated symptoms did not contribute unique information to the prediction of functioning one year later

once Time 1 teacher-rated symptoms and impairment and Time 1 parent-rated impairment were considered.  These findings suggest that, when efficiency is a priority, the use of teacher-rated symptoms in conjunction with teacher- and parent-rated impairment may provide the most useful information.

The results of the present study add to the literature supporting brief impairment ratings as a useful component of assessment in both referred and non-referred populations (Girio-Herrera, Dvorsky, & Owens, 2015; Huculak & McLennan, 2014; Sellers, Maughan, Pickles, Thapar, & Collishaw, 2015; Stringaris & Goodman, 2013).  Impairment ratings appear to reduce both false positives and false negatives and are particularly useful in community samples (Angold, Costello, Farmer, et al., 1999; Huculak & McLennan, 2014; Rapee et al., 2012).  However, despite their apparent utility, impairment ratings are not generally used in school-based mental health screening (e.g., Lane et al., 2011; White et al., 2013).  A small number of studies have examined the use of impairment ratings to determine risk in non-referred samples (Angold, Costello, Farmer, et al., 1999; Costello et al., 1999; Pickles et al., 2001; Sellers et al., 2015; Stringaris & Goodman, 2013; Wille et al., 2008) and there is some evidence of recent interest in the role of impairment ratings in universal screening for mental health difficulties (Girio-Herrera et al., 2015).  As is the case with symptom ratings, there is generally little agreement between parent and teacher ratings of impairment (Gadow et al., 2013).  Along with the findings of the present study, this suggests that parent and teacher impairment ratings contribute unique information to the assessment of risk in primary grade children.  Given that demands in home and school contexts vary greatly, it is not surprising that both informants' impairment ratings contribute to the prediction of subsequent risk.

In line with previous studies demonstrating the value of teacher symptom ratings (Dirks

et al., 2011; Drabick, Gadow, & Sprafkin, 2006; Hill et al., 2004; Shemmassian & Lee, 2012; Verhulst et al., 1994), initial teacher-rated symptoms predicted children's functioning one year later in the present sample, even after accounting for initial parent- and teacher-rated impairment and parent-rated symptoms. For example, teacher-rated symptoms in childhood and adolescence predict socioeconomic functioning in early adulthood, above and beyond parent-rated symptoms (Dirks et al., 2011). The unique information contributed by teacher-rated symptoms in addition to teacher-rated impairment provides further support for the suggestion that symptoms and impairment are related but distinct constructs (Gadow et al., 2013; Rapee et al., 2012). The present study goes a step further by demonstrating the incremental validity of teacher symptom ratings in addition to teacher and parent impairment ratings.

Consistent with previous research (Wille et al., 2008), boys were significantly more impaired than girls one year later, even after controlling for their initial levels of teacher- and parent-rated symptoms and impairment. However, boys are also more likely than girls to be false positives in mental health screening (R. Goodman et al., 2000). As a result, comprehensive follow-up assessment or ongoing monitoring are necessary once children have been identified as at-risk for mental health difficulties. This may help to ensure that the boys identified as at-risk are truly in need of mental health services.

The results presented here suggest that initial parent symptom ratings do not contribute unique information to the prediction of school-based impairment one year later once initial teacher-rated symptoms and impairment and parent-rated impairment have been considered. The low level of agreement between parent and teacher symptom ratings is well-documented (Achenbach et al., 1987; De Los Reyes et al., 2015) and previous studies have demonstrated the value of collecting both parent and teacher symptom ratings, whether the outcome is a

concurrent diagnosis or the longitudinal prediction of functioning (R. Goodman et al., 2000; Hill et al., 2004; Johnson et al., 2014; Power et al., 1998; Sowerby & Tripp, 2009; Verhulst et al., 1994).  However, not all studies support the incremental validity of parent symptom ratings.  For example, Jones and colleagues found little difference in the predictive power of models that used combined parent and teacher ratings of conduct problems versus models that used teacher ratings alone to predict future functioning (Jones, Dodge, Foster, & Nix, 2002).  Similarly, van Dulmen and Egeland (2011) found that combined parent and teacher ratings were only superior to single-informant scores when multi-informant data were aggregated based on weights that considered informant bias; otherwise, single informant data were as good as multi-informant data in terms of predicting symptoms ten years later.  Based on the findings of the present study, it appears that when a brief screening questionnaire is used in the school system, parent ratings of symptoms may be unnecessary for identifying children at risk if parent and teacher impairment ratings and teacher symptom ratings are part of the assessment.

The lack of incremental validity of parent-rated symptoms in the present study is consistent with aspects of the operations triad model proposed by De Los Reyes and colleagues to address informant discrepancies (De Los Reyes, Thomas, Goodman, & Kundey, 2013).  In brief, converging operations, in which results are perceived as more valid when there is agreement across informants, have predominated in research on multiple informants.  However, because informant discrepancies are so well-documented, there is a need for other conceptualizations, including diverging operations (in which informant discrepancies are due to meaningful differences in the child's behaviour across settings) and compensating operations (in which informant discrepancies are due to measurement error; De Los Reyes et al., 2013).  Recently, the operations triad model has been applied to the issue of incremental validity in the

assessment of children (De Los Reyes et al., 2015).  Specifically, one informant's ratings may demonstrate incremental validity if that informant observes more of the target behaviour than the other informant, an example of diverging operations.  Alternatively, if both informants observe the behaviour at a high frequency and report it reliably, it is possible that neither informant's ratings will demonstrate incremental validity due to the variance shared between them, an example of converging operations (De Los Reyes et al., 2015).  In the present study, it appears that shared variance between parent-rated symptoms and teacher-rated symptoms (an example of converging operations), and possibly between parent-rated symptoms and impairment, meant that parent-rated symptoms did not contribute uniquely to the prediction of later functioning.

More generally, parent and teacher ratings in early elementary school predicted children's functioning one year later in the present study.  The predictive value of these initial ratings argues against the "wait to fail" approach to mental health taken in a number of school systems, in which referrals for specialized services are provided only after children's difficulties begin to cause significant impairment (Baker, 2008).  On the contrary, the present findings suggest that ratings collected in early elementary school have significant prognostic value one year later.  As a result, early intervention is warranted at the point when difficulties are detected.

Finally, the results of the supplemental analysis underscore the potential effects of criterion contamination on the findings of incremental validity studies (De Los Reyes et al., 2015; Hunsley, 2003; Johnston & Murray, 2003).  In this analysis, the criterion variable more closely approximated previous approaches to examining incremental validity (e.g., Hill et al., 2004; Lochman, 1995) and was subject to criterion contamination.  In the supplemental analysis, parent-rated symptoms predicted unique variance in Time 2 risk for diagnosis but teacher-rated symptoms did not.  These findings differed from those of the primary analysis, in which a

relatively uncontaminated criterion variable was used. Although using teacher ratings of impairment one year later as the criterion variable in the primary analysis meant that criterion ratings were not based on information from any of the predictor variables, correlations among raters who observe a child in the same setting (e.g., two teachers) are consistently higher than correlations between raters who observe a child in different settings (e.g., parent and teacher; De Los Reyes et al., 2015). It is possible that this could have contributed to the lack of incremental validity of parent-rated symptoms in the primary analysis. Future studies should improve upon this by using criterion variables that are even further removed from the predictor variables, such as independent observations of behaviour (De Los Reyes et al., 2015).

### 3.5.1 Limitations and Strengths

The results presented above should be interpreted in light of some limitations that may affect the generalizability of the results. First, the findings are based on a sample of children with relatively well-educated parents compared to the Canadian average, likely all of whom spoke English. Second, a large proportion of participants were missing parent ratings. While no differences between those with and without missing data were found, it is not possible to determine whether missingness was related to the values of the missing variables themselves. For example, it is possible that parents of children experiencing more severe symptoms did not complete the questionnaires. These characteristics would reduce the level of risk in the sample but it is not clear that they would affect associations between symptom and impairment ratings and subsequent functioning. Third, respondent parents were primarily mothers and, as a result, the findings may not generalize to fathers' ratings. Because only a small proportion of the respondents in the present study were fathers, it was not possible to examine the relative utility of mothers' versus fathers' ratings. However, this warrants further investigation given that

fathers' ratings may predict more unique variance in children's functioning than mothers' ratings (Hay et al., 1999). In addition, data on teacher gender were not available and, as a result, it was not possible to examine any potential effects of teacher gender on their ratings. Fourth, incremental validity can be influenced by a variety of factors, including cultural and social context and the child's age and sex (Johnston & Murray, 2003). As a result, the findings may not generalize to samples of younger or older children or children from different social contexts (e.g., lower socioeconomic status). Further studies are also necessary to generalize these results to other screening measures and to multi-gate screening procedures (e.g., Vandeventer, 2008). Finally, the results of the present study do not provide information on appropriate cut-off scores to determine risk or ways of combining parent and teacher ratings; thus, further research addressing these issues is necessary.

The present study has several strengths. First, it is important that studies of incremental validity use "procedures, measures, and samples that reflect the realities of clinical practice" (Johnston & Murray, 2003, p. 504) and the present study design meets these criteria in terms of mental health screening. In particular, the present study used the SDQ, a brief, freely available screening questionnaire that is already in use in some school districts in North America (Huculak & McLennan, 2014; Lane et al., 2011), collected data within a school setting, and completed data collection near the beginning of the school year but after teachers had sufficient time to get to know the children. In addition, screening for at least some disorders beginning in Grade 1 appears to be a better prognostic indicator than screening at an earlier age, perhaps because the demands of Grade 1 more closely approximate the demands of later grades (Hill et al., 2004). Therefore, the present sample is young enough to provide information on early identification but old enough to have adequate prognostic value.

### 3.5.2 Implications for Practice

The findings of this study support the addition of the impact supplement, a brief measure of impairment, when screening for mental health difficulties with the SDQ in early elementary school children. Given that parent and teacher impairment ratings demonstrated incremental validity, both informants should provide impairment ratings. In addition to the empirical evidence presented here, impairment ratings may help to ensure that children with the greatest mental health needs receive services (Gordon et al., 2006). The results also suggest that elevated symptom or impairment ratings in early elementary school children are predictive of future functioning and, therefore, supports should be put in place that meet the needs of these children rather than waiting to see whether the difficulties will resolve. Finally, the results did not support the short-term prognostic value of both teacher and parent symptom ratings. Instead, when using the SDQ as a screening measure for elementary school children, the combination of teacher symptom and impairment ratings and parent impairment ratings appears to provide a balance of parsimony and prognostic value. Because the validity of impairment ratings on the SDQ has not been determined in the absence of symptom ratings, the best approach may be to have parents complete both SDQ symptom and impairment ratings but to score only the symptom ratings in the initial screening. In this way, fewer personnel resources would be needed to score the questionnaires.

### 3.5.3 Conclusion

The present study is unique in its simultaneous consideration of parent and teacher symptom and impairment ratings in screening for mental health difficulties in elementary school children. Given the growing interest in universal mental health screening, further studies are needed to inform decisions about screening approaches and to replicate the results presented

here.  Studies using other screening measures, incorporating self-report when appropriate and

examining a broader range of ages will be particularly useful.

**CHAPTER FOUR**

**Profiles of Co-Occurring Difficulties Identified through School-Based Screening**

**4.1 Abstract**

Using latent class analysis, this study identified patterns of co-occurrence among common childhood difficulties (internalizing, externalizing, inattention/hyperactivity, peer relationship problems, and reading difficulties) in 501 children ages 6 to 9. Parents and teachers rated mental health and social functioning and children completed an objective reading measure. Four latent classes were identified in the analysis of parent ratings and reading: one with externalizing, inattention/hyperactivity, peer relationship problems, and internalizing difficulties, one with inattention/hyperactivity and reading difficulties, one with internalizing and peer relationship problems, and a normative class. The analysis of teacher ratings and reading also identified four latent classes: one with externalizing and inattention/hyperactivity, one with inattention/hyperactivity and reading difficulties, one with internalizing problems, and a normative class. Children in latent classes characterized by one or more difficulties were more impaired than children in the normative latent class initially and one year later. The results highlight the need for multifaceted interventions.

**4.2 Introduction**

It is estimated that one child in eight experiences mental health difficulties within a given year (Merikangas et al., 2010).  A proactive approach, beginning with universal screening within the school system, has been recommended as a way to identify mental health difficulties early and begin providing preventive or intervention services (Dever et al., 2015; Dowdy, Ritchey, et al., 2010; Glover & Albers, 2007).  Information from early mental health screening predicts children's functioning one year later, suggesting that supports should be put in place for children who are struggling (Aitken, Martinussen, & Tannock, 2015).  However, a thorough understanding of the needs of the population is necessary in order to guide decisions about the services most likely to improve the functioning of all children (Dever et al., 2015).  Because childhood disorders tend to co-occur (Angold, Costello, & Erkanli, 1999), understanding the patterns of these co-occurring difficulties may inform the development of prevention and intervention strategies (Reinke et al., 2008).  The present study applies a relatively under-utilized person-oriented statistical approach to examine profiles of common difficulties identified through screening in a non-referred sample of children.

*4.2.1 Co-Occurring Difficulties*

Inattention/hyperactivity, internalizing problems (i.e., anxiety and depression), and externalizing problems (i.e., conduct problems and oppositional defiance) are among the most common childhood mental health concerns (Angold, Costello, & Erkanli, 1999).  These difficulties frequently co-occur in children who meet clinical diagnostic criteria as well as those with sub-clinical levels of symptoms (Angold, Costello, & Erkanli, 1999; Angold, Costello, Farmer, et al., 1999).  Children who exhibit inattention and/or hyperactivity are especially likely to experience difficulties in other areas.  They are at an increased risk for internalizing problems

relative to their peers (Cho et al., 2009; Hong et al., 2014; Leech, Larkby, Day, & Day, 2006) and they are seven times more likely than expected to exhibit externalizing problems (Waschbusch, 2002). Likewise, internalizing and externalizing problems frequently co-occur (Angold, Costello, & Erkanli, 1999; Keiley et al., 2000; Wolff & Ollendick, 2006). These findings are in line with theoretical perspectives on the etiology of mental disorders. In particular, Caspi and colleagues recently found evidence of a general psychopathology factor that may be indicative of overall risk for mental health difficulties (Caspi et al., 2014). Based on their analysis of a large, longitudinal sample of adult data, they concluded that "psychiatric disturbance tends to unfold across years of development as persistent and comorbid" (Caspi et al., 2014, p. 131). Given the high rate of overlap across symptom types, children experiencing mental health difficulties will likely require interventions that address multiple areas.

In addition to co-occurring with one another, mental health difficulties are associated with lower academic achievement. Within the area of academics, reading is particularly important because it permeates almost all curriculum areas; for example, children must read to solve mathematical word problems, read instructions in workbooks, or read for information in later grades. Inattention/hyperactivity, internalizing, and externalizing frequently co-occur with reading difficulties. For example, longitudinal studies have demonstrated the negative impact of inattention/hyperactivity on children's reading development (McGee et al., 1991; Rabiner & Coie, 2000); however, inattention appears to be more strongly associated with reading outcomes than is hyperactivity (Carroll et al., 2005). The opposite pathway, whereby reading difficulties lead to inattention/hyperactivity, has received little support (Fergusson & Horwood, 1992). In contrast, internalizing problems and reading difficulties also tend to co-occur but there is more evidence that reading difficulties contribute to the development of internalizing problems than

vice versa (Grills-Taquechel, Fletcher, Vaughn, & Stuebing, 2012; Morgan et al., 2008). Finally, children who experience reading difficulties demonstrate more externalizing behaviour on average than their peers but this association may be mediated by inattention (Carroll et al., 2005; Willcutt & Pennington, 2000). Thus, while there is substantial evidence that reading difficulties co-occur with mental health difficulties, the direction of the relationship and potential explanatory variables may vary depending on the associated mental health symptoms.

Peer relationships are an important component of children's development, so much so that children who struggle to develop friendships may experience feelings of inferiority and problems with psychological well-being (Sullivan, 1953). Indeed, there are complex bi-directional associations between peer relationship problems and a variety of mental health difficulties (Tseng et al., 2014; van Lier & Koot, 2010). Children who exhibit symptoms of inattention/hyperactivity receive lower peer social preference ratings and experience more social difficulties than other children (Bellanti & Bierman, 2000; Cho et al., 2009; Diamantopoulou et al., 2007). Similarly, externalizing problems in kindergarten and Grade 1 predict a variety of social outcomes one year later, including lower social preference ratings, more victimization, and having few friends (van Lier & Koot, 2010). In contrast, the opposite pathway has received support in the case of internalizing problems and it appears that peer relationship difficulties contribute to the development of internalizing symptoms (Hymel et al., 1990; Keiley et al., 2000; Morrow et al., 2006). Finally, children who experience academic difficulties are at an increased risk for peer relationship problems (Greenham, 1999; Wiener, 2008). These findings indicate that children's functioning in peer relationships is closely related to their mental health and academic achievement.

While the high degree of co-occurrence among common childhood mental health, academic, and social difficulties has been demonstrated, the majority of studies in this area focus on associations between pairs of disorders (Wichstrøm & Berg-Nielsen, 2014). Given the numerous associations that have been found, multivariate approaches are necessary to capture these complex interrelationships.

### 4.2.2 Person-Oriented Analyses

Most studies of co-occurring difficulties in childhood use variable-oriented analyses such as regression and structural equation modeling to examine relationships between variables. These variable-oriented methods provide important information about the prediction of outcomes and associations between the constructs studied (B. O. Muthén & Muthén, 2000); however, they also have limitations. In particular, examining interactions among several variables can become overwhelming (Cronbach, 1975). Moreover, using analyses that identify unique predictors can obscure other relevant predictors of functioning, which may also be related to the construct in important ways (Haapasalo et al., 2000).

Person-oriented statistical approaches are a useful complement to variable-oriented analyses that allow researchers to take a holistic, interactionistic perspective on childhood difficulties (Bergman & Magnusson, 1997). These analyses focus on patterns among individuals rather than associations among variables (B. O. Muthén & Muthén, 2000). Person-oriented methods can be used to detect groups of individuals who are alike in terms of the variables being studied, suggesting similarities in their functioning (Magnusson, 2003; B. O. Muthén & Muthén, 2000). The results of these analyses may contribute to our understanding of the difficulties commonly experienced by children, which may inform intervention efforts (Reinke et al., 2008).

Although cluster analysis is the most widely used person-oriented analysis (Reinke et al., 2008), model-based methods such as LCA are preferred because they generate fit indices and allow for measurement error and comparisons of alternative models (Sterba & Bauer, 2010), therefore providing a less biased analysis of patterns. In LCA, latent classes of participants are identified based on participants' pattern of scores on observed categorical variables. The goal is to determine the smallest number of latent classes that captures the patterns within the sample (B. O. Muthén & Muthén, 2000). LCA is particularly well-suited to the study of co-occurring difficulties because it provides a statistical way to identify heterogeneous groups of children who are alike in terms of their profile of difficulties (Haapasalo et al., 2000; Ostrander, Herman, Sikorski, Mascendaro, & Lambert, 2008).

Few studies have used model-based person-oriented analyses to identify patterns of co-occurring childhood difficulties. Among those that have, some have identified latent classes of children characterized by co-occurring mental health difficulties (e.g., Tolan & Henry, 1996), whereas others have identified classes characterized by difficulties in a single aspect of mental health functioning (e.g., internalizing, externalizing; Herman et al., 2012). Some studies have also applied model-based person-oriented methods to the study of academic and mental health difficulties. For example, Reinke and colleagues (2008) used LCA to identify subgroups of Grade 1 children based on academic achievement and teacher-rated behavioural problems, including aggression, inattention, and oppositionality. They identified four latent classes of boys: one with academic and behaviour problems, one with academic problems only, one with behaviour problems only, and one with no problems. The same latent classes were identified for girls, with the exception of a behaviour problems-only class. Similarly, in a recent study of Grade 3 children that used screening measures of academic and behavioural risk (based on an

overall emotional and behavioural difficulties score), three latent classes of comparable size were identified: one with minimal academic and behavioural risk, one with mainly academic risk, and one at high risk academically and moderate risk behaviourally (King et al., 2015). Collectively, these results suggest that although some children experience academic difficulties in isolation, many exhibit both academic and behavioural difficulties. However, neither of these studies considered internalizing problems separately in the creation of latent classes and, as internalizing problems are related to both externalizing and academic achievement (Masten et al., 2005), their inclusion may provide a more detailed understanding of children's profiles. In a recent exploratory study, Valdez and colleagues used LPA to identify subgroups of Grade 1 children based on internalizing, externalizing, social, and academic variables. They identified three latent classes: one that was at risk behaviourally, academically, and socially, one that was at risk academically and in peer relationships, and one that was well-adjusted (Valdez et al., 2011). Thus, co-occurring difficulties were found in the absence of latent classes with a single type of difficulty. This study was unique in its inclusion of peer acceptance as a domain of functioning; however, the authors noted that future analyses should include inattention (Valdez et al., 2011).

### 4.2.3 Current Study

No published studies were located that have applied model-based person-oriented analyses to the examination of internalizing, externalizing, inattention/hyperactivity, peer relationship problems, and reading difficulties in a community sample of children. Using person-oriented methods to examine patterns across these common childhood difficulties would provide useful information on the patterns of needs in non-referred children, which may in turn assist with prevention and early intervention planning. The results may also inform decisions about the range of areas that should be included when screening for mental health difficulties at

school. An additional consideration somewhat unique to the assessment of children's functioning is the use of multiple informants. While parents and teachers both provide useful information on children's functioning (Dirks et al., 2011), agreement between their ratings is generally low (De Los Reyes et al., 2015). Given that rater differences may be due in part to real differences in children's behaviour across settings (De Los Reyes, Henry, Tolan, & Wakschlag, 2009), it is important to gather both parent and teacher ratings when seeking to understand patterns of children's functioning.

In the present study, LCA was used to elucidate patterns of common co-occurring difficulties in the areas of internalizing, externalizing, inattention/hyperactivity, peer relationships, and reading based on screening measures in a community sample of children. Separate analyses were conducted for parent and teacher ratings to determine what patterns emerged at home and at school. The level of impairment across the latent classes was compared within and across raters both cross-sectionally and one year later to validate the latent classes and ensure they are clinically meaningful (Nylund, 2007). Based on previous studies, a large normative latent class was expected in the analysis of both parent and teacher ratings. In addition, it was predicted that inattention/hyperactivity would co-occur with difficulties in one or more areas. It was anticipated that latent classes characterized by one or more difficulties would be significantly more impaired than the normative latent class concurrently and one year later.

## 4.3 Method

### 4.3.1 Procedure

This study is part of a two-year prospective investigation of behavioural and cognitive measures of children's attention and their association with academic outcomes (for a more detailed description of the sample and procedures, see Aitken et al., 2015; Andrade & Tannock,

2013; Normand et al., 2012). Participants in Grades 1 to 3 were recruited from one suburban and one rural school board in Ontario, Canada. Data were collected in November and again one year later. Eligible participants were in a mainstream classroom, did not have sensory or physical impairments that would make them unable to complete the tasks or hear the instructions, had written parent and teacher informed consent, and provided verbal assent (Normand et al., 2012). These procedures were approved by the institutional Human Ethics Review Boards at the University of Toronto and the Hospital for Sick Children.

### 4.3.2 Sample

A total of 501 children (247 boys, 254 girls) from seven schools, ranging in age from 6 to 9 years ($M = 7.5$ years), took part in the study. The majority of the sample was White (86%) and spoke English as their primary language (97%). Twenty-five percent were enrolled in a French immersion program in which the majority of instruction is in French. Most respondent parents were mothers (90%) and they varied in terms of their highest level of education, with 2% having completed Grades 9 to 11, 6% having a high school diploma or equivalent, 10% having completed some college, 35% being college graduates, 11% having completed some university, 29% being university graduates, and 8% holding graduate degrees.

### 4.3.3 Measures

***Mental Health and Peer Relationship Problems.*** Parents (90% mothers) and teachers completed the British version of the SDQ (R. Goodman, 1997), regardless of whether the child was in French immersion. The SDQ is a 25-item behavioural screening measure based on symptoms of disorders commonly diagnosed in childhood (R. Goodman & Scott, 1999). Each item is rated on a three-point Likert-type scale (not true, somewhat true, certainly true) and items are identical across parent and teacher versions. The items are equally distributed across five

subscales, four of which were used in the present study: emotional symptoms (used as an indicator of internalizing), conduct problems (used as an indicator of externalizing), hyperactivity-inattention (used as an indicator of inattention/hyperactivity), and peer problems (used as an indicator of peer relationship problems). Parent and teacher SDQ ratings demonstrate adequate construct validity, inter-rater reliability, and internal consistency in a Canadian sample (Aitken, Martinussen, Wolfe, et al., 2015).

SDQ subscale scores can be classified as normal, borderline, or abnormal based on scoring criteria provided on the author's website (www.sdqinfo.org; see also Goodman, 1997). In a community sample, approximately 10% of children will score within the abnormal range and approximately 10% will score within the borderline range on each subscale (R. Goodman, 1997). For the purposes of the present study, parent and teacher SDQ subscale ratings were dichotomized, with children scoring in the borderline or abnormal range being classified as experiencing difficulties and children scoring within the normal range being classified as not experiencing difficulties.

*Reading Difficulties.* Children read three grade-levelled passages from the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Oral Reading Fluency task (Good & Kaminski, 2002). They read the passages out loud and the number of words read correctly in one minute was recorded, with the median of the three passages taken as the best indicator of reading ability (Good & Kaminski, 2002). Reading fluency is a good indicator of children's overall reading proficiency (Fuchs, Fuchs, Hosp, & Jenkins, 2001) and the Oral Reading Fluency measure is a reliable indicator of reading fluency, particularly when multiple passages are administered (Poncy, Skinner, & Axtell, 2005). Following Reinke and colleagues (2008), reading scores were dichotomized by classifying children falling in the bottom quartile within

their grade and language stream (French immersion vs. regular stream) as experiencing reading difficulties and other children as not experiencing reading difficulties.

**Impairment.**  The impact supplement of the SDQ (R. Goodman, 1999) was used to measure impairment at the time the SDQ was completed and again one year later.  The impact supplement consists of an additional page of questions regarding the impairment the child experiences in a variety of domains as a result of their difficulties.  The parent version contains five items and the teacher version contains three.  Each is rated on a four-point Likert-type scale (not at all, only a little, quite a lot, a great deal), with responses of "not at all" or "only a little" receiving a score of 0, "quite a lot" receiving a score of 1, and "a great deal" receiving a score of 2.  The total possible impairment score is 10 for parent ratings and 6 for teacher ratings.  Impact supplement scores have a high internal consistency for parent and teacher ratings (Stringaris & Goodman, 2013).  Both the SDQ and its impact supplement are freely available at [www.sdqinfo.org](www.sdqinfo.org).

### 4.3.4 Data Analysis

Prior to running the analyses, data cleaning and preparation were carried out.  There were 26 participants missing age data and the mean age within their grade was substituted.  There were 338 children with complete parent SDQ ratings, 8 children missing a single item, and 155 children had no parent SDQ ratings available.  There were 465 children with complete teacher SDQ ratings, 22 missing a single teacher-rated item, and 1 missing two items from different subscales (13 had no teacher SDQ ratings available).  SDQ subscale scores were prorated based on available items for participants missing individual item ratings.

Separate LCAs of parent and teacher ratings plus reading difficulties were carried out using Mplus Version 7 (L. K. Muthén & Muthén, 2012).  Each LCA included five dichotomous

variables (experiencing difficulties vs. not): internalizing, externalizing, inattention/

hyperactivity, peer relationships, and reading.  Participants with no available parent and/or

teacher SDQ ratings were excluded from the LCAs, resulting in a sample size of 346 for the

parent analysis and 488 for the teacher analysis.  The analyses were run using full information

maximum likelihood estimation (L. K. Muthén & Muthén, 2012).  Reading data were missing for

6 of the 346 participants included in the parent LCA and 10 of the 488 participants included in

the teacher LCA.  Model fit was examined for one- through five-class solutions.  Bootstrap

likelihood ratio test (BLRT), Bayesian information criterion (BIC), and entropy values were

examined for each model (Nylund, Asparouhov, & Muthén, 2007).  For the BLRT, a non-

significant value ($p \geq .05$) indicates that a model with one fewer latent classes fits the data better

(Nylund et al., 2007).  BIC is also an indicator of model fit and smaller values suggest a better

fit.  Entropy values indicate the overall accuracy of the classification and range from 0 to 1 with

values closer to 1 indicating greater accuracy (Geiser, 2013).  The BLRT was used as the

primary indicator of model fit because it identifies the correct number of latent classes more

often than do other indicators when categorical data are used (Nylund et al., 2007).  Because

theory and practical considerations, including parsimony, are important in model selection

(Collins & Lanza, 2010), the "usefulness of the latent classes in practice" (B. O. Muthén &

Muthén, 2000, p. 887) was also considered in determining the best model.

Once the best model was chosen, the latent classes were compared on sex with a cross-

tabulation and chi square test and on age with a one-way analysis of variance.  Kruskal-Wallis

tests were used to compare the latent classes on parent- and teacher-rated impairment

concurrently and one year later.  These analyses were run within and across rater types and *p*-

values < .05 were considered significant.  Follow-up comparisons were carried out with Mann-

Whitney *U* tests to determine whether there were significant differences between pairs of latent

classes. A Bonferroni correction was applied to account for multiple comparisons with a value

of $p < .0083$ considered significant. For all analyses comparing latent classes, cases were

excluded if they were missing data on the dependent variable or if they had not been assigned to

a latent class due to missing parent or teacher SDQ data.

**4.4 Results**

Because many participants were missing parent ratings, those with parent ratings ($n =$

346) were compared with those whose parents did not return the SDQ ($n = 155$). Children

missing parent ratings did not differ statistically from those with parent ratings in terms of age,

$F(1, 499) = 0.59$, $p = .44$, sex, $\chi^2(1, N = 501) = 0.22$, $p = .64$, or reading difficulties, $\chi^2(1, N =$

489) = 0.001, $p = .98$. Based on Mann-Whitney *U* tests, there were also no significant

differences in teacher-rated internalizing ($p = .77$), externalizing ($p = .13$), inattention/

hyperactivity ($p = .69$), or peer relationship problems ($p = .91$); however, participants with parent

SDQ ratings were more likely to be in French immersion, $\chi^2(1, N = 501) = 26.21$, $p < .001$, had

parents who were more educated ($p = .001$), and were rated by teachers as less impaired ($p =$

.049) than participants with missing parent SDQ subscale scores.

Table 4.1 presents the number of children experiencing internalizing, externalizing,

inattention/hyperactivity, and peer relationship difficulties based on parent and teacher ratings.

More children were generally rated as experiencing difficulties by parents than by teachers with

the exception of inattention/hyperactivity, for which teachers rated slightly more children as

experiencing difficulties than did parents. Few children experienced difficulties in the areas of

internalizing or peer relationships based on teacher ratings.

**Table 4.1**

*Proportion of Children Experiencing Difficulties*

| Domain | Parent-rated[a] | Teacher-rated[b] |
| --- | --- | --- |
| Internalizing | 17.1% | 7.6% |
| Externalizing | 17.1% | 11.7% |
| Inattention/hyperactivity | 15.9% | 18.2% |
| Peer relationships | 15.6% | 8.4% |

[a] $N = 346$; [b] $N = 488$.

### 4.4.1 Parent Ratings

*Latent Class Analysis.* Fit indices for one- through five-class solutions for parent ratings and reading are presented in Table 4.2. Although the BIC values increased from the two-class to the three-class and from the three-class to the four-class solutions, the BLRT *p*-values and the entropy values suggest that the four-class solution represents the best fit. Figure 1 shows the probability of participants in each latent class experiencing internalizing, externalizing, inattention/hyperactivity, peer relationship problems, or reading difficulties. Class 1 (4.3% of the sample) was characterized by a high probability of externalizing (1.0), inattention/hyperactivity (0.94), and peer relationship problems (1.0), a moderate probability of internalizing problems (0.57), and a low probability of reading difficulties (0.28). Thus, 100% of participants in Class 1 exhibited externalizing and peer relationship problems, 94% exhibited inattention/hyperactivity, 57% experienced internalizing problems, and 28% experienced reading difficulties. Based on these probabilities, Class 1 was labeled Multiple Difficulties. Participants in Class 2 (6.9% of the sample) had a high probability of inattention/hyperactivity (1.0) and reading difficulties (0.77) and zero probability of peer relationship problems. Class 2 was labeled ADHD symptoms and Reading Difficulties. Class 3 (4.9% of the sample) had a high

probability of internalizing problems (1.0) and peer relationship problems (1.0) and a very low

probability of externalizing problems (0.02).  Class 3 was labeled Internalizing and Social

Difficulties.  Finally, Class 4 (83.8% of the sample) had a low probability of internalizing (0.10),

externalizing (0.12), inattention/hyperactivity (0.05), peer relationship problems (0.08), and

reading difficulties (0.21) and was labeled the Normative class.

**Table 4.2**

*Model Fit Indices for Latent Class Analyses*

| Solution | BIC | BLRT $p$-value | Entropy |
|---|---|---|---|
| Parent[a] | | | |
| 1 class | 1644.25 | – | – |
| 2 classes | 1598.96 | <.001 | 0.85 |
| 3 classes | 1612.03 | <.001 | 0.92 |
| **4 classes** | **1629.12** | **.01** | **0.95** |
| 5 classes | 1655.35 | .09 | 0.98 |
| Teacher[b] | | | |
| 1 class | 1924.20 | – | – |
| 2 classes | 1789.99 | .00 | 0.74 |
| 3 classes | 1812.46 | .01 | 0.87 |
| **4 classes** | **1836.09** | **<.001** | **0.95** |
| 5 classes | 1864.06 | 1.00 | 0.89 |

*Note.* Bold indicates the best-fitting model.  BIC = Bayesian

information criterion; BLRT = bootstrap likelihood ratio test.

[a]$n = 346$; [b]$n = 488$.

The LCAs were re-run accounting for the nesting of children within classrooms and the

pattern of latent classes and the values of the fit indices did not change, although there were

small differences in the standard errors.  As a result, the original analyses were retained to allow

the use of the BLRT.  Because the entropy value for the four-class solution is above .80, it is

appropriate to assign participants to their most likely latent classes and compare the classes in

further analyses (Clark & Muthén, 2009).  There were no significant age, $F(3, 342) = 0.87$, $p =$

.46, or sex differences, $\chi^2(3, N = 346) = 0.65$, $p = .89$, across the latent classes.

**Figure 4.1**

*Probability of Difficulties by Latent Class Based on Parent Ratings*
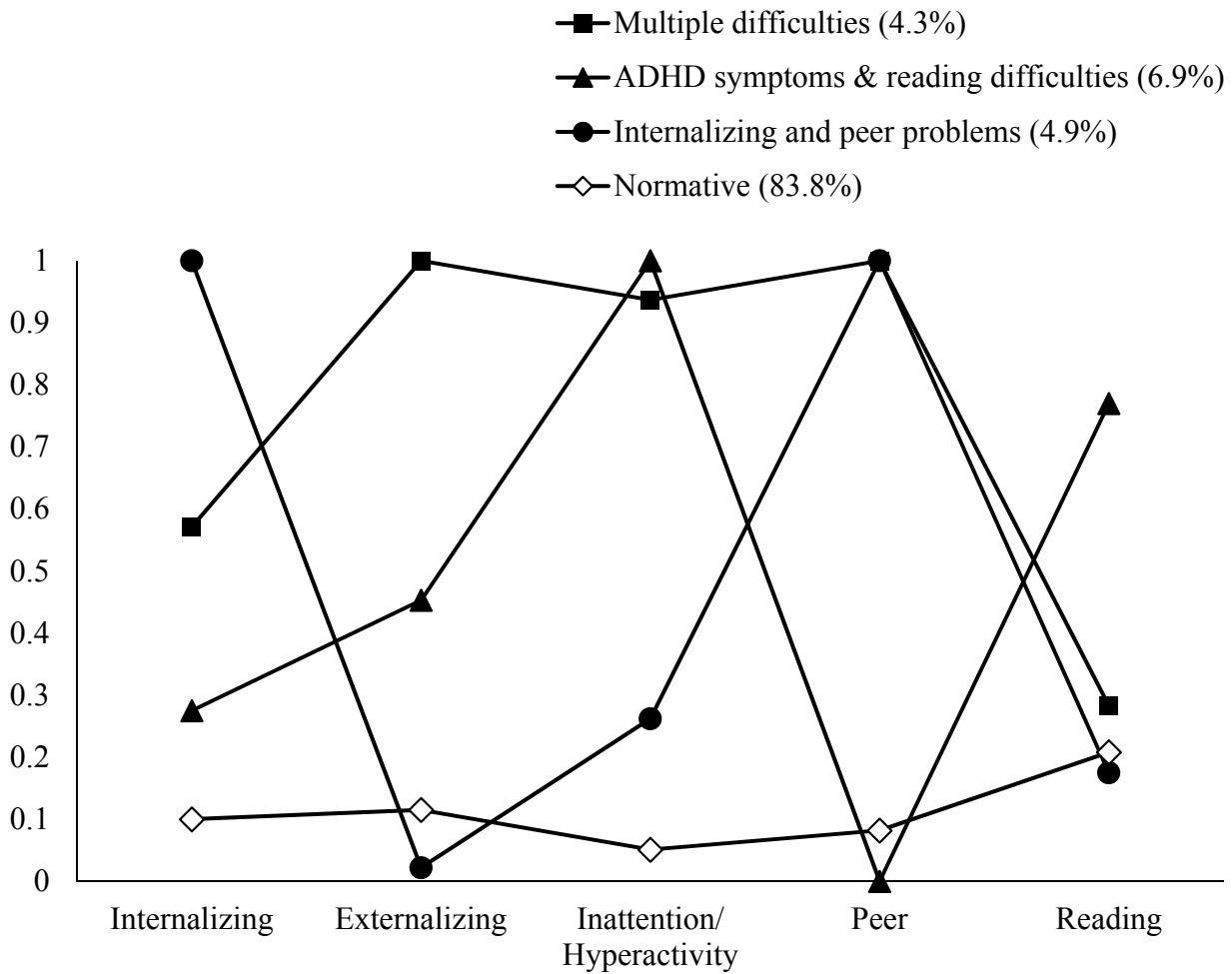


*Concurrent Impairment across Latent Classes.*  Means and standard deviations for

parent- and teacher-rated impairment are presented for each latent class in Table 4.3.  The latent

classes generated based on parent ratings differed significantly in terms of both parent- and teacher-rated impairment ($p < .001$). Post hoc tests examining parent-rated impairment indicated that participants in the Multiple Difficulties, ADHD Symptoms and Reading Difficulties, and the Internalizing and Social Difficulties latent classes were all significantly more impaired than the Normative latent class (all $p$s $< .001$). No other differences were significant. When teacher-rated impairment was examined in post hoc tests, the Multiple Difficulties and the ADHD Symptoms and Reading Difficulties latent classes were significantly more impaired than the Normative class ($p$s $< .001$). Participants in the Internalizing and Social Difficulties latent class were not significantly more impaired than those in the Normative latent class ($p = 0.03$, Bonferroni corrected $p$-value set at $< .0083$) and no other comparisons were significant.

**Table 4.3**

*Means and Standard Deviations for Impairment Ratings by Latent Class*

| Class label | % of participants | Parent-rated impairment | | Teacher-rated impairment | |
|---|---|---|---|---|---|
| | | Time 1[a] | Time 2[b] | Time 1[c] | Time 2[d] |
| Parent-derived latent classes | | | | | |
| Multiple difficulties | 4.3 | **2.60 (2.44)** | **2.38 (1.94)** | **1.13 (1.64)** | **1.62 (2.18)** |
| ADHD symptoms & reading | 6.9 | **1.29 (1.73)** | **1.68 (2.03)** | **1.04 (1.27)** | **1.63 (1.64)** |
| Internalizing & social | 4.9 | **1.38 (2.22)** | **1.46 (2.22)** | 0.63 (1.09) | **0.77 (1.48)** |
| Normative | 83.8 | 0.21 (0.77) | 0.27 (0.76) | 0.22 (0.70) | 0.28 (0.85) |
| Teacher-derived latent classes | | | | | |
| ADHD symptoms & reading | 10.0 | **0.78 (1.49)** | **1.13 (1.61)** | **0.95 (1.08)** | **1.47 (1.87)** |
| Self-regulation difficulties | 73.8 | **1.86 (2.30)** | **2.00 (2.07)** | **1.69 (1.61)** | **1.71 (1.90)** |
| Internalizing problems | 11.7 | 0.50 (0.94) | **1.38 (2.00)** | **0.43 (0.85)** | **0.38 (0.74)** |
| Normative | 4.5 | 0.20 (0.77) | 0.20 (0.65) | 0.05 (0.27) | 0.15 (0.51) |

*Note.* Values in bold indicate a significant difference from the Normative class.

[a] $n = 341$ for parent classes, $n = 336$ for teacher classes; [b] $n = 280$ for parent classes, $n = 353$ for teacher classes;

[c] $n = 339$ for parent classes, $n = 487$ for teacher classes; [d] $n = 331$ for parent classes, $n = 461$ for teacher classes.

***Latent Class Membership Predicting Future Impairment.***  Significant differences

among the latent classes were found for both parent- and teacher-rated impairment one year later

(*p*s < .001; see Table 4.3 for means and standard deviations).  Post hoc tests indicated that

participants in the Multiple Difficulties, ADHD Symptoms and Reading Difficulties, and the

Internalizing and Social Difficulties latent classes were each significantly more impaired than

participants in the Normative latent class based on both parent (*p*s < .007) and teacher ratings (*p*s

< .003).  No other differences were significant.

### 4.4.2 Teacher Ratings

***Latent Class Analysis.***  Table 4.2 presents the fit indices for one- through five-class

solutions for teacher ratings and reading.  A four-class solution represented the data best based

on the significant BLRT *p*-value, the entropy value, and the extraction of meaningful latent

classes.  The probability of participants in each latent class experiencing internalizing,

externalizing, inattention/hyperactivity, peer relationship problems, and reading difficulties is

presented in Figure 2.  Children in Class 1 (10% of the sample) had a high probability of

inattention/hyperactivity (1.0) in the absence of any externalizing problems (0.0) and a moderate

probability of reading difficulties (0.59).  Class 1 was labeled ADHD Symptoms and Reading

Difficulties.  Participants in Class 2 (73.8% of the sample) had a low probability of internalizing

(< .01), externalizing (0.01), inattention/hyperactivity (0.03), peer relationship problems (0.02),

and reading difficulties (0.19).  This class was labeled the Normative class.  Class 3 (11.7% of

the sample) was characterized by a high probability of externalizing problems (1.0) and

inattention/hyperactivity (0.76).  Class 3 was labeled Self-Regulation Difficulties.  Finally,

participants in Class 4 (4.5% of the sample) had a high probability of internalizing problems

(1.0), a moderate probability of reading difficulties (0.45), and zero probabilities of

inattention/hyperactivity and externalizing. It was labeled the Internalizing Problems class.

Similar results were obtained when the LCAs were re-run accounting for the nesting of children

within classrooms.

**Figure 4.2**

*Probability of Difficulties by Latent Class Based on Teacher Ratings*
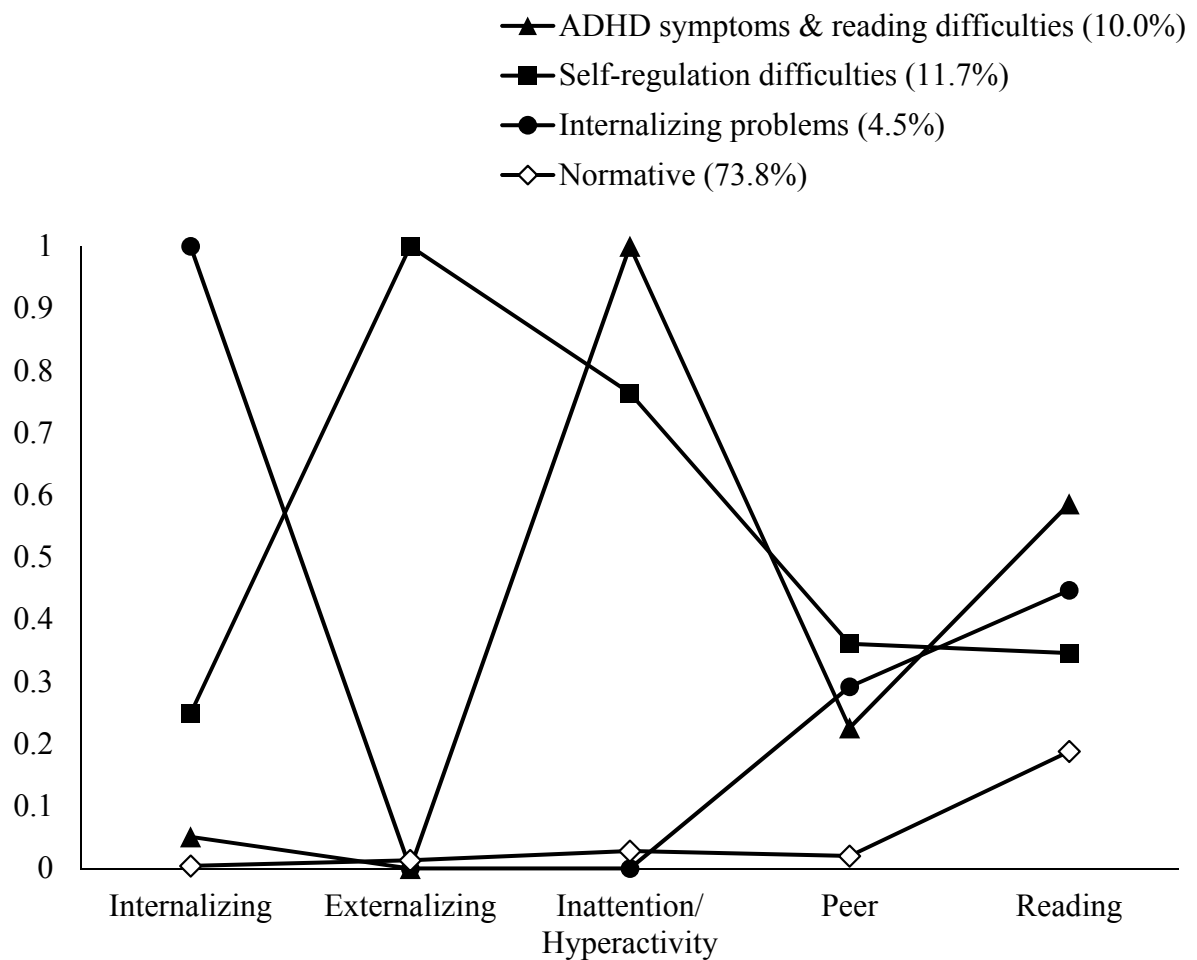


The entropy value for the four-class solution was large enough for participants to be

assigned to their most likely latent class in order to compare the classes (Clark & Muthén, 2009).

There were no significant age differences across the four latent classes, $F(3, 484) = 1.53$, $p = .21$;

however, there were significant sex differences, $\chi^2(3, N = 488) = 27.45$, $p < .001$. Post hoc

comparisons (Bonferroni-corrected $p$-value = 0.006) indicated that boys were more likely than expected to be in the ADHD Symptoms and Reading Difficulties class, $\chi^2(1) = 17.81$, $p < .001$, and less likely than expected to be in the Normative class, $\chi^2(1) = 21.07$, $p < .001$.  The inverse was true for girls, who were more likely than expected to be in the Normative class, $\chi^2(1) = 21.07$, $p < .001$, and less likely than expected to be in the ADHD Symptoms and Reading Difficulties class, $\chi^2(1) = 17.81$, $p < .001$.  No other differences were significant.

*Concurrent Impairment across Latent Classes.*  Mean impairment ratings for the latent classes are presented in Table 4.3.  The latent classes differed significantly in terms of concurrent teacher-rated impairment ($p < .001$) and all post hoc comparisons between classes were significant.  Specifically, participants in the ADHD Symptoms and Reading Difficulties, Self-Regulation Difficulties, and Internalizing Problems latent classes were all significantly more impaired than participants in the Normative latent class (all $ps < .001$).  Participants in the Self-Regulation Difficulties class were the most impaired, significantly more so than those in the ADHD Symptoms and Reading Difficulties class ($p = .001$) or the Internalizing Problems class ($p < .001$).  Children in the ADHD Symptoms and Reading Difficulties class were also more impaired than children in the Internalizing Problems class ($p = .0081$).

The latent classes also differed significantly in concurrent parent-rated impairment ($p < .001$; see Table 4.3).  Post hoc tests indicated that participants in the ADHD Symptoms and Reading Difficulties class and the Self-Regulation difficulties class were rated as significantly more impaired by parents than those in the Normative class ($ps < .001$); however, participants in the Internalizing Problems class were not significantly more impaired than those in the Normative class ($p = .04$, Bonferroni corrected $p$-value set at $< .0083$).  No other comparisons were significant.

***Latent Class Membership Predicting Future Impairment.*** The latent classes generated based on teacher ratings also differed significantly on both teacher- and parent-rated impairment one year later ($ps < .001$; see Table 4.3 for means). Pairwise comparisons indicated that participants in the ADHD Symptoms and Reading Difficulties, Self-Regulation Difficulties, and Internalizing Problems classes were significantly more impaired than participants in the Normative class based on teacher ($ps \leq .001$) and parent impairment ratings ($ps \leq .002$). No other differences among the latent classes were significant.

## 4.5 Discussion

The present study applied person-oriented analysis to a range of common childhood difficulties, including internalizing, externalizing, inattention/hyperactivity, peer relationship problems, and reading difficulties. The goal was to determine what profiles of co-occurring difficulties are present in a community sample of early elementary school children when data are gathered in a way that simulates broad-based screening. As expected, large Normative latent classes of participants were identified in both parent and teacher LCAs in which children had a low probability of difficulties in any of the five areas. Latent classes of children experiencing co-occurring difficulties in two or more areas were also identified in the analyses of parent and teacher ratings. All participants in latent classes characterized by difficulties in one or more areas were significantly more impaired than participants in the Normative latent class based on cross-sectional analyses using the same rater (parent or teacher) and longitudinal analyses both within and across rater types. These findings suggest that the latent classes are clinically meaningful and indicative of children's level of need both concurrently and one year later.

This study is one of only a small number that have used person-oriented analyses to capture patterns of common childhood difficulties. In general, the entropy values indicated that

the LCAs of both parent and teacher ratings fit the data well, classifying children with 95% accuracy. The size of the latent classes was also as expected for a community sample of children, with a large proportion of children (84% in the parent analysis, 74% in the teacher analysis) falling within a "Normative" class that had a low probability of any difficulties. The remaining latent classes provide novel information on the predominant profiles of common childhood difficulties.

Inattention/hyperactivity is known to co-occur with a variety of social, emotional, behavioural, and academic difficulties (Cho et al., 2009; Diamantopoulou et al., 2005; Rabiner & Coie, 2000; Waschbusch, 2002). In the present study, two latent classes in each of the parent and teacher analyses were characterized by a high probability of inattention/hyperactivity, suggesting that it manifests in a variety of ways and with various co-occurring difficulties. In line with previous variable-oriented studies (McGee et al., 1991; Rabiner & Coie, 2000), both the parent and teacher LCAs identified a latent class of children with a high probability of inattention/hyperactivity and a high or moderate probability of reading difficulties. Similarly, latent classes in which inattention/hyperactivity and externalizing behaviour co-occurred were identified based on both parent and teacher ratings. These findings provide further support for the high degree of overlap between inattention/hyperactivity and conduct or oppositional problems (Waschbusch, 2002). The results also support previous suggestions that the presence of externalizing problems contributes to the development of other difficulties in children who are hyperactive and/or inattentive (Andrade & Tannock, 2013; Becker, Luebbe, & Langberg, 2012; Carlson, Tamm, & Gaub, 1997). In particular, in the analysis of parent ratings, internalizing problems and peer relationship problems co-occurred with inattention/hyperactivity, but only within a latent class of children who also had a high probability of externalizing behaviour.

Thus, it is important to assess for difficulties in other areas when children present with hyperactivity and/or inattention.  Interventions for these children will need to be multifaceted and tailored to their specific pattern of difficulties.

The co-occurrence of internalizing and social difficulties has been well-documented and there is evidence of a causal pathway, whereby social difficulties contribute to the development of internalizing problems (Hymel et al., 1990; Keiley et al., 2000; Morrow et al., 2006).  More recently, Van Lier and Koot (2010) demonstrated a developmental cascade from externalizing behaviour in kindergarten to later peer relationship problems, which in turn contributed to internalizing problems in Grade 4.  In line with this developmental cascade pathway, the present LCA of parent ratings identified a latent class of children experiencing externalizing and peer problems (as well as inattention/hyperactivity), in which just over half of the participants also experienced internalizing problems (Class 1).  However, peer relationship problems and internalizing were also present without externalizing problems in Class 3 in the parent LCA and this latent class was of comparable size to Class 1.  These findings suggest that while some children may experience a cascade effect from externalizing to social difficulties and internalizing, others develop internalizing and social difficulties in the absence of conduct problems or oppositional behaviour.  Similar findings were reported in a sample of Grade 4 children, in which those who were socially withdrawn and rejected experienced more self-reported depression than other children, whether or not they were aggressive (Boivin et al., 1994).

Previous studies have also demonstrated associations between peer relationship problems and mental health difficulties (Diamantopoulou et al., 2007; Hymel et al., 1990; van Lier & Koot, 2010).  However, peer relationship problems did not characterize any of the latent classes

in the analysis of teacher ratings in the present study, although the Normative class did have the lowest probability of difficulties in this area. This may be in part because teachers identified a relatively small percentage of children as experiencing peer relationship problems. Therefore, it appears that when teachers identify peer relationship problems, they co-occur at similar rates across other difficulties, such as inattention/hyperactivity, externalizing, and internalizing.

Results of person-oriented studies are informative not only for the patterns they identify but also for the patterns they do not identify (referred to as "white spots"; e.g., Bergman & Trost, 2006). These white spots indicate that a particular pattern is unlikely to occur. In the present study, peer relationship problems and reading difficulties did not co-occur in any of the latent classes, despite previous research suggesting that children with learning disabilities experience social difficulties (Greenham, 1999; Wiener, 2008). Given that most previous studies have examined samples of older elementary school children (e.g., Greenham, 1999), it is possible that the relatively young age of the children in the present study explains the lack of co-occurring reading and peer relationship difficulties. In addition, the analysis of teacher ratings did not identify any latent classes of children in which both externalizing and internalizing problems were present. The presence of an internalizing latent class in the analysis of teacher ratings argues against the interpretation that teachers are unable to identify children experiencing internalizing problems. Instead, it is possible that teachers are less sensitive than parents to internalizing problems in the presence of externalizing behaviour. Finally, with the exception of the Internalizing Problems class in the analysis of teacher ratings (which also had a moderate probability of reading difficulties), no "pure" difficulties latent classes were identified in the analysis of parent or teacher ratings. This provides further support for the high rates of co-

occurring academic, social, emotional, and behavioural difficulties, even in non-referred elementary school children (Angold, Costello, Farmer, et al., 1999).

Because LCA uses statistical methods to identify patterns of individuals, it is important to ensure that the latent classes derived are meaningful (Nylund, 2007).  As a way of validating the latent classes, the level of impairment was compared across the classes based on both parent and teacher ratings.  Within rater types, children in all of the latent classes characterized by high probabilities of one or more difficulties were significantly more impaired than children in the Normative class.  When concurrent impairment levels across raters were examined (e.g., parent-derived latent classes predicting teacher ratings of impairment), latent classes characterized by inattention/hyperactivity, with or without externalizing problems and reading difficulties, were associated with significantly higher impairment ratings than those of the Normative class.  However, children in latent classes characterized by internalizing problems (the Internalizing and Social Difficulties class in the parent LCA and the Internalizing Problems class in the teacher LCA) were not significantly more impaired than children in the Normative class in these concurrent cross-rater analyses.  Interestingly, when impairment ratings were examined one year later, all children in latent classes characterized by one or more difficulties, including those in the Internalizing and Social Difficulties class in the parent LCA and the Internalizing Problems class in the teacher LCA, were significantly more impaired than children in the Normative class, both within and across rater types.  These results demonstrate the clinical utility of the latent classes generated based on parent and teacher ratings and reading performance and their prognostic value one year later.  They also suggest that although children with primarily internalizing presentations may not experience impairment across settings upon initial assessment, over time their level of impairment becomes more pervasive.  Finally, the impairment ratings highlight the

importance of providing intervention for children experiencing social, emotional, behavioural, or academic difficulties as their functioning is significantly affected and this tends to persist over time.

### 4.5.1 Limitations

The results of the present study should be interpreted considering some limitations. First, the use of dichotomous variables has been criticized as it can result in a loss of information (MacCallum, Zhang, Preacher, & Rucker, 2002). However, some authors have shown that the use of continuous variables does not increase predictive power above and beyond dichotomous predictors (Farrington & Loeber, 2000; Jones et al., 2002). Dichotomous variables were used in the present study due to the strong positive skew of the SDQ scores and the interpretability of dichotomous variables in terms of risk (Farrington & Loeber, 2000). Second, this study used a single, brief measure of internalizing, externalizing, inattention/hyperactivity, and peer relationship problems. The use of a more comprehensive measure or measures of these constructs may be more informative in terms of clinical presentations; however, the present study was intended to demonstrate patterns identified in a non-referred sample of children in a screening situation. Third, parent ratings were missing for a large number of children and those with missing parent ratings were rated as more impaired by teachers. As a result, the impairment level of the latent classes generated based on parent ratings may have been underestimated. Fourth, there is some evidence that different profiles of co-occurring difficulties may be present in boys and girls (Reinke et al., 2008); however, the sample size in the present study was not large enough to allow for separate analyses for boys and girls. Similarly, the sample size did not allow for separate analyses to be run for older and younger children. Finally, the education levels of parents of participating children in the present study were above the Canadian average

and, therefore, the results may not generalize to populations with more environmental risk factors such as lower socioeconomic status.

The SDQ also includes a strength-based prosocial behaviour subscale which was not used in the present analysis due to the focus on common childhood difficulties. However, ratings on this prosocial subscale partially account for the association between hyperactivity/impulsivity and peer relationship problems (Andrade & Tannock, 2013). In addition, prosociality may buffer the effects of moderate externalizing problems on social functioning (Andrade, Browne, & Tannock, 2014). As a result, ratings on the prosocial behaviour subscale may be useful in intervention planning.

### 4.5.2 Conclusion

A significant proportion of non-referred early elementary school children experience difficulties in two or more important aspects of mental health, social, or academic functioning at home or at school. Using LCA, the present study demonstrated that the most common profiles of difficulties identified based on parent screening ratings include: 1) externalizing, inattention/hyperactivity, peer relationship problems, and internalizing; 2) inattention/hyperactivity and reading difficulties; and 3) internalizing and peer relationship problems. A parallel analysis of teacher ratings identified the following profiles of difficulties: 1) externalizing and inattention/hyperactivity; 2) inattention/hyperactivity and reading difficulties; and 3) internalizing problems. These findings provide further evidence of the high degree of co-occurrence among common childhood difficulties and support the need for a broad-based approach that includes social and academic functioning when screening for mental health difficulties in schools. They also provide information on the types of difficulties school systems must be prepared to address following universal screening of young children and highlight the

need for interventions that target difficulties in multiple areas.  Children in all of the latent

classes characterized by one or more difficulties were more impaired than children in the

normative latent class both cross-sectionally and one year later.  Further research is necessary to

determine whether providing appropriate intervention may interrupt these negative trajectories

and improve children's functioning at home and at school.

**CHAPTER FIVE**

**Discussion and Implications**

**5.1 Summary of Key Findings**

The overall goal of this dissertation was to contribute to knowledge in the area of screening for mental health difficulties.  Based on the three studies presented herein, several important findings emerged.  First, the study presented in Chapter 2 supports the reliability and validity of the SDQ, a brief and freely available screening questionnaire, as a measure of Canadian children's mental health and social functioning based on parent and teacher ratings.  Second, the primary analysis presented in Chapter 3 indicates that parent and teacher ratings of impairment and teacher ratings of symptoms each contribute unique information when screening for mental health difficulties in a school context.  Although a different pattern of results was obtained in a supplemental analysis, the present discussion will focus on the results of the primary analysis because they are subject to less criterion contamination.  The results of that study also provide further evidence that early mental health difficulties are not transient and are associated with poorer functioning one year later.  Finally, the study presented in Chapter 4 suggests that elementary school children often struggle in more than one aspect of functioning, which may include internalizing, externalizing, inattention/hyperactivity, peer relationship problems, or reading difficulties.  Because the relation of these results to previous research as well as their limitations have been discussed in the preceding chapters, this chapter will focus on situating the findings within a broader theoretical context, practical implications, and directions for future research.

**5.2 Theoretical Context**

The results of the studies examining incremental validity of multiple informants (Chapter 3) and patterns of co-occurring difficulties (Chapter 4) fit within a broader theoretical context in terms of assessing and understanding mental health.  In particular, the results can be related to

theoretical models that account for informant differences, the co-occurrence of difficulties, and associations between competence and mental health, each of which will be discussed in greater detail below.

### 5.2.1. Informant Differences

In the present dissertation, teacher ratings of children's symptoms were a more useful predictor of children's school-based functioning one year later than were parent ratings of symptoms. Parent symptom ratings did not demonstrate incremental validity above and beyond teacher symptom ratings and parent and teacher impairment ratings. There are three potential explanations for these results based on existing theories of multi-informant ratings, two of which are based on the operations triad model. The operations triad model proposes that agreement across informant ratings is indicative of converging operations (in which both informants observe and rate the behaviour in a similar way), whereas differences in informant ratings are due to diverging operations (actual differences in behaviour across settings) or compensating operations (measurement error; De Los Reyes et al., 2013, 2015). Based on this model, ratings from one informant will not demonstrate incremental validity if the target behaviour was not displayed in the setting in which the informant observes the child (diverging operations; De Los Reyes et al., 2015). However, this seems unlikely in the present study given that parent ratings identified more children as at-risk for internalizing, externalizing, and peer relationship problems than did teacher ratings. A second possible explanation based on the same model is that shared variance among parent and teacher symptom ratings resulted in parent symptom ratings being a nonsignificant predictor (converging operations). Thus, if this explanation is correct, parent symptom ratings overlap to such an extent with teacher symptom ratings that parent ratings do not contribute additional information to predicting functioning one year later. Although parent

and teacher symptom ratings were only moderately correlated in the present sample, shared variance between parent symptom and impairment ratings may also have contributed to the lack of significance for parent symptom ratings.

As a potential third explanation for the results, Dirks and colleagues have suggested that "symptoms occurring in different situations or settings, or as perceived by different informants, may constitute distinct phenotypes" (Dirks et al., 2012, p. 559), which is similar to earlier recommendations that disorders be conceptualized as "informant-specific" (Offord et al., 1996). In the case of the present study, given that the primary goal was to identify children who may benefit from school-based mental health services, it is logical that ratings of symptoms in the school environment were a better predictor of later school-based impairment than were symptoms displayed in the home environment. It is interesting, then, that parent ratings of impairment continued to predict future school-based impairment, even once earlier teacher ratings of impairment were considered. Given that the vast majority of research on multiple informants has focused on symptom ratings (De Los Reyes et al., 2015; Dirks et al., 2012), impairment ratings may represent a fruitful area for further exploration in terms of the operations triad model and informant-specific disorders.

### 5.2.2 Factors Underlying Co-Occurrence

In accounting for the co-occurrence of difficulties in different areas, the present dissertation drew on two complementary theoretical models: 1) the multiple deficit model (Pennington, 2006), in which comorbidity occurs because two or more disorders share a common risk factor; and 2) the developmental cascades model (e.g., Masten & Cicchetti, 2010), in which comorbidity is the result of difficulties in one domain affecting other aspects of functioning. These two models can also be understood in the context of recent findings suggesting that a

general psychopathology factor ("p") exists, which is indicative of individuals' overall susceptibility to mental health difficulties (Caspi et al., 2014). Although p has been studied primarily in adult populations, poorer childhood self-control and cognitive functioning are associated with higher p scores in adulthood (Caspi et al., 2014), suggesting that overall risk for mental health difficulties may also be present in childhood. However, as Caspi and colleagues (2014) note, the p factor is not necessarily causal and may instead be the result of developmental cascades over the lifespan. Thus, in this way, the notion of a general risk for mental health difficulties is similar to the multiple deficit model but does not preclude developmental cascade processes, which may also be operating. The results of the LCA presented in Chapter 4 are agnostic regarding the reasons common childhood difficulties co-occur; however, they provide further evidence that many children who struggle in one domain will also need support in other aspects of their functioning. It seems plausible that there are multiple explanations for the co-occurrence of childhood difficulties which may differ depending upon the nature of the difficulties under consideration.

### 5.2.3 Mental Health and Competence

Competence in important domains such as academic achievement and peer relationships is often considered alongside mental health functioning in research on developmental cascades (Masten et al., 2005; Murray-Close et al., 2010; van Lier & Koot, 2010). Similarly, low competence in these important areas has been theorized to contribute to depressive symptoms by causing children to internalize negative self-perceptions (Cole, Jacquez, & Maschman, 2001; Cole, Maxwell, & Martin, 1997). The results of the LCA presented in Chapter 4 provide further evidence that functioning in peer relationships and academic achievement is related to aspects of mental health such as externalizing, internalizing, and inattention/hyperactivity. Therefore, it

would be beneficial for studies of school-based mental health to consider these competence domains alongside emotional and behavioural functioning (Atkins, Hoagwood, Kutash, & Seidman, 2010).  Doing so would further our knowledge of the processes that may be contributing to or exacerbating children's mental health difficulties as well as the way mental health may affect academic and social functioning.  These competence domains may also hold important information for intervention, whether in the form of protective factors that can be drawn upon or additional areas in which children may require support.

**5.3 Practical Implications**

The results of the present dissertation have a number of practical implications in terms of mental health screening, particularly within the school system.  These implications are discussed below, along with the broader context of school-based screening and early intervention for mental health difficulties.

*5.3.1 Use of the SDQ*

The SDQ has been used in a number of research- and school-based screening projects (R. Goodman et al., 2004, 2000; Lane et al., 2011; White et al., 2013); however, based on data from the United States, it is only currently used in approximately 9% of schools that screen for mental health difficulties (Bruhn et al., 2014).  Given that many schools use school- or board-developed screening measures (Bruhn et al., 2014), it is important to promote the use of valid and reliable screening measures such as the SDQ rather than measures that may have questionable psychometric properties (Bruhn et al., 2014).  Although data are not available on the frequency with which the SDQ is being used in Canada, a number of Canadian researchers have used the measure in recent publications (Andrade & Tannock, 2013; Huculak & McLennan, 2014; Law et al., 2014), suggesting that it may be becoming more widely known.

The results of the CFA presented in Chapter 2 provide preliminary support for the reliability and validity of the British version of the parent and teacher SDQ when used with Canadian elementary school children. However, psychometric properties are only one of several important characteristics of screening questionnaires and other factors, such as usability and appropriateness must also be considered (Glover & Albers, 2007). In terms of usability, the SDQ is a brief measure that is generally well-received by parents (R. Goodman & Scott, 1999), and accurately differentiates between high risk and low-risk children (R. Goodman & Scott, 1999; R. Goodman, 1997; Klasen et al., 2000). The SDQ is also appropriate for school-based mental health screening because it is designed for use with elementary school children and measures a range of common mental health difficulties as well as social functioning and strengths in prosocial behaviour. In addition, the SDQ is relatively brief, which minimizes the amount of time required for teachers to complete it. Because the SDQ is freely available in the public domain, it is also cost-effective as a screening tool. Thus, Canadian researchers and clinicians may wish to use parent and teacher ratings on the SDQ as a way to capture aspects of mental health and social functioning.

### 5.3.2 Choice of Informants

The results of the primary incremental validity analyses (Chapter 3) suggest that parent symptom ratings do not contribute to better prediction of future functioning once teacher symptom and impairment ratings and parent impairment ratings have been considered. While this may appear to suggest that parent symptom ratings are unnecessary when screening for mental health difficulties in the school system, caution is warranted before drawing this conclusion. This is particularly true given that the results of the LCA (Chapter 4) indicate that several meaningful patterns of co-occurring difficulties are detected in parents' ratings of

children on the SDQ.  Thus, while parent symptom ratings may not contribute unique information to the broad prediction of risk, they do provide information that is useful in terms of understanding the nature of children's difficulties.  Moreover, having parents provide symptom ratings would not require many more resources than having them complete impairment ratings alone.  Therefore, an appropriate solution could involve asking parents to complete symptom and impairment ratings but scoring and using only teacher symptom and impairment ratings and parent impairment ratings for the initial screening.  Parent symptom ratings could then be examined, along with teacher symptom ratings, for children who are deemed at-risk based on the initial screening in order to better understand the pattern of difficulties they are experiencing.  Further research is necessary to replicate the results regarding incremental validity and to determine the acceptability of this proposed approach among parents and teachers.

### 5.3.3 Impairment Ratings

The results presented in Chapter 3 also support the inclusion of brief parent and teacher impairment ratings when screening for mental health difficulties in children.  In summary, both parent and teacher ratings of impairment predicted unique variance in school-based impairment one year later, above and beyond parent and teacher ratings of symptoms.  These findings fit well with the recent interest in using impairment ratings to assess risk for mental health difficulties in non-referred children and youth (DuPaul et al., 2014; Girio-Herrera et al., 2015; Sellers et al., 2015; Stringaris & Goodman, 2013).  The use of impairment ratings in mental health screening is also consistent with clinical diagnostic practices (APA, 2013; Stringaris & Goodman, 2013).  In addition, there is a linear association between impairment and symptom severity in children and being above or below a symptom-based diagnostic threshold is not associated with further differences in concurrent impairment (Pickles et al., 2001).  Therefore, there is no reason that

impairment ratings should be reserved only for comprehensive psychological assessments.

When considered alongside current diagnostic practices and evidence of the association between

symptoms and impairment, the results of this dissertation suggest that impairment ratings are

useful when screening for mental health difficulties.  These findings have broad implications for

mental health screening, whether it is carried out in schools or in other settings, such as

physicians' offices.

### 5.3.4 Broad-Based Screening and Multifaceted Interventions

The results of the LCA presented in Chapter 4 provide further evidence that peer

relationship problems and reading difficulties co-occur with mental health difficulties,

particularly as rated by parents.  However, the primary focus of research and practice in the area

of school-based screening has been on academic achievement (Bruhn et al., 2014; Cook et al.,

2010).  The findings presented here provide further evidence that assessing mental health and

social functioning in conjunction with academic achievement (e.g., reading) would provide a

more comprehensive picture of children's functioning (Flanagan et al., 2003; Lane, Oakes, &

Menzies, 2010).  Given that numerous mediational associations have been demonstrated among

academic achievement, mental health, and peer relationships (Andrade & Tannock, 2013;

Herman et al., 2007; Morgan et al., 2008), this information would be valuable for intervention

planning.  In addition, mental health difficulties can have a negative impact on children's

academic functioning and, as a result, addressing mental health issues is important for school

functioning (Atkins et al., 2010; Esch et al., 2014).  Recently, *aimsweb,* a commercial, online

screening and progress monitoring system, began providing behaviour screening as an add-on to

the original literacy and numeracy screening measures (Pearson, 2013).  However, other reliable

and valid screening measures, such as the SDQ, could also be used in conjunction with academic

screeners to identify children who may be at risk at school due to academic difficulties, mental health difficulties, or both.

The co-occurrence of peer relationship problems, academic difficulties, internalizing, externalizing, and inattention/hyperactivity also has implications for intervention. In particular, some studies have reported that co-occurring difficulties reduce the effectiveness of interventions (Hagan-Burke et al., 2011; Halldorsdottir & Ollendick, 2014; Ollendick, Jarrett, Grills-Taquechel, Hovey, & Wolff, 2008; although the majority of studies reviewed in Ollendick et al., 2008, did not find that that comorbidity had a negative effect on treatment outcomes). In addition, interventions that target a specific area of difficulty may not translate into improvements in other aspects of functioning. For example, reading interventions do not lead to improvements in externalizing, internalizing, or social skills (Roberts, Solis, Ciullo, McKenna, & Vaughn, 2015). Thus, it has been suggested that "creative treatment programs that address the complexities of comorbid disorders will be needed" (Ollendick et al., 2008, p. 1466). However, caution is warranted given that multi-component school-based mental health interventions are not always superior to more focused interventions, perhaps because intervention efforts become diluted and are not implemented with high fidelity (Weare & Nind, 2011). Encouragingly, there is some evidence that interventions targeting underlying risk factors may result in improvements in multiple domains. For example, Chase and Eyberg (2008) found that parent-child interaction therapy led to improvements in terms of separation anxiety and oppositional defiance. Therefore, the judicious use of multi-component interventions as well as interventions that target shared underlying risk factors are promising ways to address the challenge of co-occurring difficulties.

In addition to addressing difficulties across multiple aspects of functioning, interventions may need to change aspects of children's environment in order to be successful (Herman et al., 2009). For example, Mikami and Normand (2015) propose that the limited effectiveness of interventions for peer relationship problems in children with ADHD is due to the fact that only social skills and behaviour within the child with ADHD are targeted, without changing the social context in which the child functions. As a result, they propose that classroom-based interventions that increase peer acceptance of children with ADHD while also addressing challenging behaviours may be more effective (Mikami & Normand, 2015). Schools are an ideal setting in which to provide ecologically valid interventions that address both target behaviours as well as the context in which children are functioning (Stephan et al., 2007).

### 5.3.5 Additional Considerations

In discussing the practical implications of the present dissertation, it is important to consider the broader challenges associated with universal screening and school-based mental health services. First, screening alone is not likely to be beneficial and it may have negative consequences if appropriate interventions are not provided (Sayal et al., 2010; Tymms & Merrell, 2006). Second, there are legal and practical challenges to implementing universal screening programs for mental health (Kirby & Keon, 2006). In particular, Kirby and Keon (2006) note the need for informed consent and maintaining confidentiality as well as schools being prepared to meet the needs of children identified as at-risk. Another legal and ethical issue that warrants thorough consideration is how the results of school-based screening are shared. For example, Tymms and Merrell (2006) speculate that providing feedback to parents may be more useful than providing feedback to schools, although they did not test this hypothesis. Legally, however, it may be necessary to first provide parents with the results of the screening

and disclose the results to the school only after obtaining parental consent.  As a result, it may be necessary for outside agencies or professionals such as school psychologists to carry out the screening, rather than teachers or school principals.  Third, the various funding envelopes that may be involved in school-based mental health in Canada represent an additional challenge.  For example, given that the screening and potentially the intervention services would be provided through the school, it is logical to assume that, in Ontario, the funding would come from the Ministry of Education.  However, it is possible that school board budgets may not be able to absorb this extra cost.  Indeed, the Mental Health Commission of Canada's report on school-based mental health identified funding as the greatest challenge to program implementation (School-Based Mental Health and Substance Abuse Consortium, 2013).  As a result, other branches of government, such as the Ministry of Children and Youth Services, which funds children's mental health services in Ontario (www.children.gov.on.ca), may need to be involved.  A related issue is the threshold at which children would be eligible for services within the school system.  At present, mental health services in Canadian schools are generally only provided to students experiencing significant levels of difficulty (School-Based Mental Health and Substance Abuse Consortium, 2013); therefore, it is unclear whether children experiencing elevated but sub-clinical levels of difficulty would be eligible for mental health services under the current model.  In summary, significant challenges remain in terms of initiating mental health screening and early intervention within the school system.  Although the present dissertation addresses a number of important empirical issues, additional research is also necessary in order to guide the implementation of universal screening for mental health within the school system.

**5.4 Directions for Future Research**

The results of the three studies presented here provide a foundation for future research on school-based mental health screening. Important areas for further exploration include collecting Canadian normative data, examining potential differences in mothers' and fathers' ratings, investigations of older children and adolescents that include self-report, the use of other assessment tools, issues of cost and feasibility, and pilot studies of school-based screening in conjunction with early intervention. These will be discussed in more detail below.

*5.4.1 Canadian Normative Data*

At present, Canadian normative data are not available for the SDQ. Thus, in order to use the SDQ with Canadian children, it is necessary to refer to either British normative data (as was the case in the present studies, which used the British version of the SDQ) or to use the United States version of the SDQ (which has slightly different wording) and its corresponding normative data (Bourdon, Goodman, Rae, Simpson, & Koretz, 2005). While the use of normative data from other countries is not ideal, it is common practice in Canada. For example, the Behavior Assessment System for Children – 2nd Edition (BASC-2; Reynolds & Kamphaus, 2004) is widely used in Canada and does not include Canadian children in its normative sample. A better solution would be to obtain normative data on the SDQ for Canadian children and youth. This would involve collecting data from a large, nationally representative sample of Canadian children and adolescents. For example, in Bourdon and colleagues' (2005) study, data from 10,367 4- to 17-year-olds were collected through the United States Census Bureau's annual National Health Interview Survey. Cut-offs could then be established for each subscale as well as for the total difficulties score such that 80% of children fall within the normal range, followed by a band of 10% who fall within the borderline range, and finally, the most extreme 10% falling

within the abnormal range (R. Goodman, 1997). In the United States normative sample, the cut-offs were found to be similar to those of the British sample, although a cut-off of 1 or 2 points lower was set for the total difficulties score to reflect the slightly lower average scores of the children in the Unites States sample (Bourdon et al., 2005). Thus, it would be important to have more precise information with which to compare the functioning of Canadian children. Future studies should also examine the discriminant validity of the SDQ in a Canadian sample, including sensitivity, specificity, positive predictive value and negative predictive value.

Along these lines, the findings reported in Chapter 3 provide evidence that parent and teacher impairment ratings have prognostic value above and beyond teacher ratings of symptoms. However, further research is needed to determine the best way to combine impairment and symptom ratings when screening for mental health difficulties. Given that the goal of screening is to identify all children at risk, including those with sub-clinical difficulties, an "or" rule may be appropriate. For example, children who meet the cut-off for risk either based on symptoms or impairment could be referred for further assessment or monitored for a period of time.

### 5.4.2 Mothers versus Fathers as Informants

Studies that request parent ratings of children's emotional and behavioural functioning often receive reports primarily from mothers (Owens et al., 2015; Power et al., 1998; Stringaris & Goodman, 2013) and this was also the case in the present study. However, just as there is limited agreement between parent and teacher ratings, correlations between mothers' and fathers' ratings are only moderate ($r = .48$ for internalizing problems and $r = .58$ for externalizing problems based on a meta-analysis; De Los Reyes et al., 2015). Mothers and fathers may also provide different information on children's functioning. For example, a recent study using the

SDQ found that fathers rated children higher than mothers on the conduct problems, hyperactivity-inattention, and emotional symptoms subscales (Chiorri, Hall, Casely-Hayford, & Malmberg, 2015); however, other studies have reported that mothers tend to provide higher ratings than fathers (Christensen, Margolin, & Sullaway, 1992) or that mothers and fathers provide similar ratings (Hay et al., 1999). Of note, Hay and colleagues (1999) found that, when considered together, fathers' ratings but not mothers' ratings predicted unique variance in teacher-rated functioning later in childhood. Therefore, fathers' ratings are a potentially useful source of information that should be investigated in future studies of incremental validity.

### 5.4.3 Older Children and Adolescents

Given the focus on early school-based intervention for mental health difficulties, the present dissertation used a sample of primary grade children (ages 6 to 9 years); however, the SDQ was designed for use with a wider range of children and adolescents ages 4 to 17. Some studies that have examined a broader age range of children have found that alternative models of the SDQ's factor structure fit better than the five-factor model. In particular, two recent studies have found support for a bifactor model with two correlated global factors (Internalizing, comprising the emotional symptoms and peer problems subscales, and Externalizing, comprising the hyperactivity-inattention and conduct problems subscales; Caci et al., 2015; Kóbor et al., 2013). Therefore, although the study presented in Chapter 2 indicated that the five-factor model fit the data well with a relatively narrow age range of children, future studies may need to examine global factors in order to obtain a good model fit.

In addition, there is some evidence that inter-rater agreement is higher for younger children than for older children (Achenbach et al., 1987). Therefore, it would be important to examine the incremental validity of parent and teacher ratings for older children and adolescents

in future studies. A self-report version of the SDQ is also available for youth ages 11 to 17, which would be useful to incorporate in future studies of incremental validity, given the relative ease with which self-report ratings could be collected in a school setting.

It has also been suggested that the structure of psychopathology may change across development. Specifically, childhood psychopathology may be more general (Nottelmann & Jensen, 1995) but may become more defined and specific as children develop further in terms of personality and neurology (Zahn-Waxler et al., 2000). As a result, it is possible that different patterns of co-occurring difficulties may characterize older children and adolescents. Therefore, studies that apply person-oriented methods to examine patterns of co-occurring academic, social, and mental health difficulties in older samples may be informative (see Okado & Bierman, 2014 for an example that examined latent classes based on internalizing and externalizing problems).

### 5.4.4 Other Measures of Childhood Difficulties

The study presented in Chapter 4 provides information on patterns of co-occurring difficulties that may be detected in elementary school children based on a brief screening measure. However, richer information on these patterns may be obtained by using more comprehensive measures of internalizing, externalizing, inattention/hyperactivity, peer relationship problems, and academic functioning. In terms of internalizing, externalizing, and inattention/hyperactivity, this could include broad band measures such as the CBCL and Teacher Report Form (Achenbach, 1991) or narrow band measures such as the Children's Depression Inventory-2 (Kovacs, 2011), the Multidimensional Anxiety Scale for Children – 2nd Edition (March, 2013), and the Conners-3 (Conners, 2008). Likewise, sociometric methods that use children's peers as informants are considered the "gold standard" (e.g., Hoza et al., 2005, p. 412) as indicators of social functioning. Therefore, using approaches such as peer nominations (e.g.,

most liked, least liked) or peer ratings may provide more valid information on children's peer relationship functioning than parent and teacher ratings. Finally, it would be informative to include more comprehensive measures of academic achievement that include written expression and mathematics in addition to reading.

### 5.4.5 Cost and Feasibility

Currently, little is known about the cost of school-based mental health screening and its potential benefits in relation to these costs (Dodge et al., 2015; Kuo et al., 2009). The present set of studies provides guidance in terms of the selection of measures and informants that will allow research to move forward in this area. Examining cost and cost-effectiveness is a logical next step once the incremental validity of an assessment practice has been demonstrated (Hunsley, 2003; Yates & Taub, 2003). In the case of school-based mental health screening, the cost of the screening (in terms of physical and human resources) should be weighed against the potential benefits. Benefits can be measured in terms of monetary (e.g., lower cost to society, increased earning potential for the individual) and non-monetary outcomes (e.g., mental health, well-being; Yates & Taub, 2003). In the case of screening, further work remains to be done initially in determining the cost of procedures such as using teacher symptom ratings and parent and teacher impairment ratings on the SDQ to screen for mental health problems. Then, the benefits of this screening, in conjunction with evidence-based intervention, could be examined (Yates & Taub, 2003 provide a description of the steps to take to determine the cost-benefit of psychological assessment, which could also be applied to mental health screening).

### 5.4.6 Pilot Studies Including Early Intervention

Given that mental health screening is only likely to be beneficial to the extent that it facilitates early intervention, it is important to conduct future studies that examine the

effectiveness of intervention delivered as a result of identification through screening. Dodge and colleagues (2015) provide a methodologically rigorous example of one such study, in which children were randomly assigned to intervention or control conditions after being identified as at-risk for conduct problems. The intervention delivered in that study is best characterized as Tier 3 given its intensity and comprehensiveness. However, other approaches that lend themselves better to a tiered system are possible, such as those that include school-wide (Tier 1) components, as well as selective intervention that is less intensive than Tier 3 (i.e., Tier 2). Based on a review of evidence-based school mental health programs, Weare and Nind (2011) concluded that the most effective programs are characterized by: a balance of universal and targeted intervention, a focus on early intervention with continued work with older students, skill teaching, the use of positive mental health approaches, and changes across the broader system of parents, school, and community. By conducting pilot studies of school-based screening in conjunction with interventions that meet these criteria, data could be gathered on the overall effectiveness of early identification and intervention for mental health difficulties through the school system.

**5.5 Conclusion**

A significant proportion of Canadian children experience mental health difficulties and delivering services within the school system has been proposed as a way to increase their access to supports. Universal screening for mental health difficulties may help to identify the children who would benefit most from these services. This dissertation presented three studies that were collectively intended to increase our knowledge regarding school-based mental health screening. The results indicate that when a reliable and valid screening measure is used, teacher ratings of symptoms and impairment and parent ratings of impairment are predictive of future functioning

at school, suggesting that children who experience mental health difficulties should receive intervention services as early as possible. Many of the children who are at risk experience difficulties in multiple areas and, although parent symptom ratings may not add to the prediction of risk over time, they do provide rich information on the co-occurring difficulties that may need to be addressed through intervention. These findings represent an important foundational step that may guide future work in the area of mental health screening in terms of selecting measures and informants and anticipating potential patterns of need in young children. However, many challenges and questions remain, making mental health screening and early intervention an exciting avenue for further exploration, which is greatly needed before programs of this nature are implemented.

# References

Achenbach, T. M. (1991). *Manual for the teacher's report form and 1991 profile*. Burlington, VT: University of Vermont, Department of Psychiatry.

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: implications of cross-informant correlations for situational specificity. *Psychological Bulletin, 101*(2), 213–232. doi:10.1037/0033-2909.101.2.213

Aitken, M., Martinussen, R., & Tannock, R. (2015). Incremental validity of parent and teacher symptom and impairment ratings when screening for emotional and behavioural disorders, Manuscript under review.

Aitken, M., Martinussen, R., Wolfe, R. G., & Tannock, R. (2015). Factor structure of the Strengths and Difficulties Questionnaire in a Canadian elementary school sample. *Assessment for Effective Intervention, 40*(3), 155–165. doi:10.1177/1534508414560347

Aitkin, M., Anderson, D., & Hinde, J. (1981). Statistical modeling of data on teaching styles. *Journal of the Royal Statistical Society, Series A, 144*, 419–461.

Albers, C. A., Glover, T. A., & Kratochwill, T. R. (2007). Introduction to the special issue: How can universal screening enhance educational and mental health outcomes? *Journal of School Psychology, 45*(2), 113–116. doi:10.1016/j.jsp.2006.12.002

Allison, P. D. (2012). Regression for count data. In *Logistic Regression Using SAS: Theory and Applications* (pp. 265–290). Cary, NC: SAS Institute, Inc.

Al-Yagon, M., & Margalit, M. (2006). Loneliness, sense of coherence and perception of teachers as a secure base among children with reading difficulties. *European Journal of Special Needs Education, 21*(1), 21–37. doi:10.1080/08856250500268619

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders*

(4th Ed.,TR). Washington, DC: Author.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th Ed.). Washington, DC: Author.

Andrade, B. F., Browne, D. T., & Tannock, R. (2014). Prosocial skills may be necessary for better peer functioning in children with symptoms of disruptive behavior disorders. *PeerJ, 2*, e487. doi:10.7717/peerj.487

Andrade, B. F., & Tannock, R. (2013). The direct effects of inattention and hyperactivity/impulsivity on peer problems and mediating roles of prosocial and conduct problem behaviors in a community sample of children. *Journal of Attention Disorders, 17*(8), 670–80. doi:10.1177/1087054712437580

Angold, A., Costello, E. J., & Erkanli, A. (1999). Comorbidity. *Journal of Child Psychology and Psychiatry, 40*(1), 57–87.

Angold, A., Costello, E. J., Farmer, E. M. Z., Burns, B. J., & Erkanli, A. (1999). Impaired but undiagnosed. *Journal of the American Academy of Child and Adolescent Psychiatry, 38*(2), 129–37. doi:10.1097/00004583-199902000-00011

Atkins, M. S., Hoagwood, K. E., Kutash, K., & Seidman, E. (2010). Toward the integration of education and mental health in schools. *Administration and Policy in Mental Health and Mental Health Services Research, 37*, 40–47. doi:10.1007/s10488-010-0299-7

Baker, J. A. (2008). Assessing school risk and protective factors. In B. Doll & J. A. Cummings (Eds.), *Transforming school mental health services: Population-based approaches to promoting the competency and wellness of children* (pp. 43–68). Thousand Oaks, CA: Corwin Press.

Baldwin, J. S., & Dadds, M. R. (2008). Examining alternative explanations of the covariation of

ADHD and anxiety symptoms in children: a community study. *Journal of Abnormal Child Psychology*, *36*(1), 67–79. doi:10.1007/s10802-007-9160-1

Barriga, A. Q., Doran, J. W., Newell, S. B., Morrison, E. M., Barbetti, V., & Robbins, B. D. (2002). Relationships between problem behaviors and academic achievement in adolescents: The unique role of attention problems. *Journal of Emotional and Behavioral Disorders*, *10*(4), 233–240.

Becker, S. P., Luebbe, A. M., & Langberg, J. M. (2012). Co-occurring mental health problems and peer functioning among youth with attention-deficit/hyperactivity disorder: A review and recommendations for future research. *Clinical Child and Family Psychology Review*, *15*(4), 279–302. doi:10.1007/s10567-012-0122-y

Bellanti, C. J., & Bierman, K. L. (2000). Disentangling the impact of low cognitive ability and inattention on social behavior and peer relationships. *Journal of Clinical Child & Adolescent Psychology*, *29*(1), 66–75. doi:10.1207/S15374424jccp2901

Benningfield, M. M., & Stephan, S. H. (2015). Integrating mental health into schools to support student success. *Child and Adolescent Psychiatric Clinics of North America*, *24*(2), xv–xvii. doi:10.1016/j.chc.2014.12.005

Bergman, L. R., & Magnusson, D. (1997). A person-oriented approach in research on developmental psychopathology. *Development and Psychopathology*, *9*(2), 291–319.

Bergman, L. R., & Trost, K. (2006). The person-oriented versus the variable-oriented approach: Are they complementary, opposites, or exploring different worlds? *Merrill-Palmer Quarterly*, *52*(3), 601–632. doi:10.1353/mpq.2006.0023

Bird, H. R., Gould, M. S., Rubio-Stipec, M., Staghezza, B. M., & Canino, G. (1991). Screening for childhood psychopathology in the community using the Child Behavior Checklist.

*Journal of the American Academy of Child & Adolescent Psychiatry*, *30*(1), 116–123.

Bird, H. R., Gould, M. S., & Staghezza, B. M. (1993). Patterns of diagnostic comorbidity in a community sample of children aged 9 through 16 years. *Journal of the American Academy of Child and Adolescent Psychiatry*, *32*(2), 361–368. doi:http://dx.doi.org/10.1097/00004583-199303000-00018

Boak, A., Hamilton, H. A., Adlaf, E. A., Beitchman, J., Wolfe, D., & Mann, R. E. (2014). *The Mental Health and Well-Being of Ontario Students, 1991-2013: Detailed OSDUHS Findings*. Retrieved from http://www.camh.ca

Boivin, M., Poulin, F., & Vitaro, F. (1994). Depressed mood and peer rejection in childhood. *Development and Psychopathology*, *6*(03), 483. doi:10.1017/S0954579400006064

Bourdon, K. H., Goodman, R., Rae, D. S., Simpson, G., & Koretz, D. S. (2005). The Strengths and Difficulties Questionnaire: U.S. normative data and psychometric properties. *Journal of the American Academy of Child and Adolescent Psychiatry*, *44*(6), 557–564. doi:10.1097/01.chi.0000159157.57075.c8

Breton, J. J., Bergeron, L., Valla, J. P., Berthiaume, C., Gaudet, N., Lambert, J., … Lépine, S. (1999). Quebec child mental health survey: prevalence of DSM-III-R mental health disorders. *Journal of Child Psychology and Psychiatry*, *40*(3), 375–384. doi:10.1017/s0021963098003618

Bringewatt, E. H., & Gershoff, E. T. (2010). Falling through the cracks: Gaps and barriers in the mental health system for America's disadvantaged children. *Children and Youth Services Review*, *32*(10), 1291–1299. doi:10.1016/j.childyouth.2010.04.021

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford.

Bruhn, A. L., Woods-Groves, S., & Huddle, S. (2014). A preliminary investigation of emotional

and behavioral screening practices in K–12 schools. *Education and Treatment of Children*, *37*(4), 611–634. doi:10.1353/etc.2014.0039

Bryan, T. H. (1976). Peer popularity of learning disabled children: A replication. *Journal of Learning Disabilities*, *9*(5), 49–53.

Caci, H., Morin, A. J. S., & Tran, A. (2015). Investigation of a bifactor model of the Strengths and Difficulties Questionnaire. *European Child & Adolescent Psychiatry*, (Advance online publication). doi:10.1007/s00787-015-0679-3

Canadian Institute for Health Information. (2015). *Care for Children and Youth With Mental Disorders*. Ottawa, ON. Retrieved from http://www.cihi.ca

Carlson, C. L., Tamm, L., & Gaub, M. (1997). Gender differences in children with ADHD, ODD, and co-occurring ADHD/ODD identified in a school population. *Journal of the American Academy of Child and Adolescent Psychiatry*, *36*(12), 1706–1714. doi:10.1097/00004583-199712000-00019

Carroll, J. M., Maughan, B., Goodman, R., & Meltzer, H. (2005). Literacy difficulties and psychiatric disorders: evidence for comorbidity. *Journal of Child Psychology and Psychiatry*, *46*(5), 524–32. doi:10.1111/j.1469-7610.2004.00366.x

Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., … Moffitt, T. E. (2014). The p factor: One general psychopathology factor in the structure of psychiatric disorders? *Clinical Psychological Science*, *2*(2), 119–137. doi:10.1177/2167702613497473

Chafouleas, S. M., Kilgus, S. P., & Wallach, N. (2010). Ethical dilemmas in school-based behavioral screening. *Assessment for Effective Intervention*, *35*(4), 245–252. doi:10.1177/1534508410379002

Chapman, J. W., Tunmer, W. E., & Prochnow, J. E. (2000). Early reading-related skills and performance, reading self-concept, and the development of academic self-concept: A longitudinal study. *Journal of Educational Psychology*, *92*(4), 703–708. doi:10.1037/0022-0663.92.4.703

Chase, R. M., & Eyberg, S. M. (2008). Clinical presentation and treatment outcome for children with comorbid externalizing and internalizing symptoms. *Journal of Anxiety Disorders*, *22*(2), 273–282. doi:10.1016/j.janxdis.2007.03.006

Chatterji, P., Caffray, C. M., Crowe, M., Freeman, L., & Jensen, P. (2004). Cost assessment of a school-based mental health screening and treatment program in New York City. *Mental Health Services Research*, *6*(3), 155–166.

Children's Mental Health Ontario. (2015). *Pre-Budget Submission 2015*. Retrieved from http://www.kidsmentalhealth.ca

Chiorri, C., Hall, J., Casely-Hayford, J., & Malmberg, L.-E. (2015). Evaluating measurement invariance between parents using the Strengths and Difficulties Questionnaire (SDQ). *Assessment*, *Advance on*. doi:10.1177/1073191114568301

Cho, S. C., Kim, B. N., Kim, J. W., Rohde, L. A., Hwang, J. W., Chungh, D. S., … Kim, H. W. (2009). Full syndrome and subthreshold attention-deficit/hyperactivity disorder in a Korean community sample: comorbidity and temperament findings. *European Child & Adolescent Psychiatry*, *18*(7), 447–57. doi:10.1007/s00787-009-0755-7

Christensen, A., Margolin, G., & Sullaway, M. (1992). Interparental agreement on child behavior problems. *Psychological Assessment*, *4*(4), 419–425. doi:10.1037/1040-3590.4.4.419

Clark, S. L., & Muthén, B. O. (2009). *Relating Latent Class Analyis Results to Variables not Included in the Analysis*.

Cohen, M. A., & Piquero, A. R. (2009). New evidence on the monetary value of saving a high risk youth. *Journal of Quantitative Criminology*, *25*(1), 25–49. doi:10.1007/s10940-008-9057-3

Cole, D. A., Jacquez, F. M., & Maschman, T. L. (2001). Social origins of depressive cognitions: A longitudinal study of self-perceived competence in children. *Cognitive Therapy and Research*, *25*(4), 377–395. doi:10.1023/A:1005582419077

Cole, D. A., Maxwell, S. E., & Martin, J. M. (1997). Reflected self-appraisals: Strength and structure of the relation of teacher, peer, and parent ratings to children's self-perceived competencies. *Journal of Educational Psychology*, *89*(1), 55–70. doi:10.1037/0022-0663.89.1.55

Cole, D. A., Truglio, R., & Peeke, L. (1997). Relation between symptoms of anxiety and depression in children: a multitrait-multimethod-multigroup assessment. *Journal of Consulting and Clinical Psychology*, *65*(1), 110–119. doi:10.1037/0022-006X.65.1.110

Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences.* Hoboken, NJ: Wiley.

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis*. Hillsdale, NJ: Erlbaum.

Conners, C. K. (2008). *Conners-3*. North Tonawanda, NY: Multi-Health Systems.

Conti-Ramsden, G., Mok, P. L. H., Pickles, A., & Durkin, K. (2013). Adolescents with a history of specific language impairment (SLI): Strengths and difficulties in social, emotional and behavioral functioning. *Research in Developmental Disabilities*, *34*(11), 4161–4169. doi:10.1016/j.ridd.2013.08.043

Cook, C. R., Volpe, R. J., & Livanis, A. (2010). Constructing a roadmap for future universal screening research beyond academics. *Assessment for Effective Intervention*, *35*(4), 197–

205. doi:10.1177/1534508410379842

Copeland, W. E., Adair, C. E., Smetanin, P., Stiff, D., Briante, C., Colman, I., … Angold, A. (2013). Diagnostic transitions from childhood to adolescence to early adulthood. *Journal of Child Psychology and Psychiatry*, *54*(7), 791–799. doi:10.1111/jcpp.12062

Costello, E. J., Angold, A., & Keeler, G. P. (1999). Adolescent outcomes of childhood disorders: The consequences of severity and impairment. *Journal of the American Academy of Child and Adolescent Psychiatry*, *38*(2), 121–128. doi:10.1097/00004583-199902000-00010

Costello, E. J., Mustillo, S., Erkanli, A., Keeler, G., & Angold, A. (2003). Prevalence and development of psychiatric disorders in childhood and adolescence. *Archives of General Psychiatry*, *60*(8), 837–44. doi:10.1001/archpsyc.60.8.837

Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist*, *30*(2), 116–127. doi:10.1037/h0076829

Dadds, M. R., Holland, D. E., Laurens, K. R., Mullins, M., Barrett, P. M., & Spence, S. H. (1999). Early intervention and prevention of anxiety disorders in children: results at 2-year follow-up. *Journal of Consulting and Clinical Psychology*, *67*(1), 145–150.

Dally, K. (2006). The influence of phonological processing and inattentive behavior on reading acquisition. *Journal of Educational Psychology*, *98*(2), 420–437. doi:10.1037/0022-0663.98.2.420

De Los Reyes, A., Augenstein, T. M., Wang, M., Thomas, S. A., Deborah, A. G., Burgers, D. E., … Rabinowitz, J. (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychological Bulletin*, Advance online publication. doi:10.1037.a0038498

De Los Reyes, A., Henry, D. B., Tolan, P. H., & Wakschlag, L. S. (2009). Linking informant

discrepancies to observed variations in young children's disruptive behavior. *Journal of Abnormal Child Psychology*, *37*(5), 637–642. doi:10.1007/s10802-009-9307-3

De Los Reyes, A., Thomas, S. A., Goodman, K. L., & Kundey, S. M. (2013). Principles underlying the use of multiple informants' reports. *Annual Review of Clinical Psychology*, *9*, 123–49. doi:10.1146/annurev-clinpsy-050212-185617

Deater-Deckard, K. (2001). Annotation: Recent research examining the role of peer relationships in the development of psychopathology. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *42*(5), 565–579. doi:10.1111/1469-7610.00753

Dever, B. V, Dowdy, E., Raines, T. C., & Carnazzo, K. (2015). Stability and change of behavioral and emotional screening scores. *Psychology in the Schools*, *52*(6), 618–629. doi:10.1002/pits.21825

Dever, B. V, Raines, T. C., & Barclay, C. M. (2012). Chasing the unicorn: Practical implementation of universal screening for behavioral and emotional risk. *School Psychology Forum*, *6*(4), 108–118.

Diamantopoulou, S., Henricsson, L., & Rydell, A. M. (2005). ADHD symptoms and peer relations of children in a community sample: Examining associated problems, self-perceptions, and gender differences. *International Journal of Behavioral Development*, *29*(5), 388–398. doi:10.1177/01650250500172756

Diamantopoulou, S., Rydell, A. M., Thorell, L. B., & Bohlin, G. (2007). Impact of executive functioning and symptoms of attention deficit hyperactivity disorder on children's peer relations and school performance. *Developmental Neuropsychology*, *32*(1), 521–42. doi:10.1080/87565640701360981

Dickey, W. C., & Blumberg, S. J. (2004). Revisiting the factor structure of the Strengths and

Difficulties Questionnaire: United States, 2001. *Journal of the American Academy of Child and Adolescent Psychiatry*, *43*(9), 1159–1167. doi:10.1097/01.chi.0000132808.36708.a9

Dirks, M. A., Boyle, M. H., & Georgiades, K. (2011). Psychological symptoms in youth and later socioeconomic functioning: do associations vary by informant? *Journal of Clinical Child and Adolescent Psychology, 40*(1), 10–22. doi:10.1080/15374416.2011.533403

Dirks, M. A., De Los Reyes, A., Briggs-Gowan, M., Cella, D., & Wakschlag, L. S. (2012). Annual research review: Embracing not erasing contextual variability in children's behavior - Theory and utility in the selection and use of methods and informants in developmental psychopathology. *Journal of Child Psychology and Psychiatry*, *53*(5), 558–574. doi:10.1111/j.1469-7610.2012.02537.x

Dodge, K. A., Bierman, K. L., Coie, J. D., Greenberg, M. T., Lochman, J. E., McMahon, R. J., & Pinderhughes, E. E. (2015). Impact of early intervention on psychopathology, crime, and well-being at age 25. *American Journal of Psychiatry*, *172*(1), 59–70. doi:10.1176/appi.ajp.2014.13060786

Dodge, K. A., & Pettit, G. S. (2003). A biopsychosocial model of the development of chronic conduct problems in adolescence. *Developmental Psychology*, *39*(2), 349–371. doi:10.1037/0012-1649.39.2.349

Dowdy, E., Doane, K., Eklund, K., & Dever, B. V. (2013). A comparison of teacher nomination and screening to identify behavioral and emotional risk within a sample of underrepresented students. *Journal of Emotional and Behavioral Disorders*. doi:10.1177/1063426611417627

Dowdy, E., Furlong, M., Eklund, K., Saeki, E., & Ritchey, K. (2010). Screening for mental health and wellness: Current school-based practices and emerging possibilities. In B. Doll & W. Pfohl (Eds.), *Handbook of youth prevention science* (pp. 70–95). New York: Routledge.

Dowdy, E., & Kim, E. (2012). Choosing informants when conducting a universal screening for behavioral and emotional risk. *School Psychology Forum*, *6*(4), 1–10.

Dowdy, E., Nylund-Gibson, K., Felix, E. D., Morovati, D., Carnazzo, K. W., & Dever, B. V. (2014). Long-rerm stability of screening for behavioral and emotional risk. *Educational and Psychological Measurement*, *74*(3), 453–472. doi:10.1177/0013164413513460

Dowdy, E., Ritchey, K., & Kamphaus, R. W. (2010). School-based screening: a population-based approach to inform and monitor children's mental health needs. *School Mental Health*, *2*, 166–176. doi:10.1007/s12310-010-9036-3

Drabick, D. A., Gadow, K. D., & Loney, J. (2008). Co-occurring ODD and GAD symptom groups: source-specific syndromes and cross-informant comorbidity. *Journal of Clinical Child and Adolescent Psychology*, *37*(2), 314–26. doi:10.1080/15374410801955862

Drabick, D. A., Gadow, K. D., & Sprafkin, J. (2006). Co-occurrence of conduct disorder and depression in a clinic-based sample of boys with ADHD. *Journal of Child Psychology and Psychiatry*, *47*(8), 766–74. doi:10.1111/j.1469-7610.2006.01625.x

DuPaul, G. J., Reid, R., Anastopoulos, A. D., & Power, T. J. (2014). Assessing ADHD symptomatic behaviors and functional impairment in school settings: Impact of student and teacher characteristics. *School Psychology Quarterly*, *29*(4), 409–421. doi:10.1037/spq0000095

Durlak, J. A., & Wells, A. M. (1998). Evaluation of indicated preventive intervention (secondary prevention) mental health programs for children and adolescents. *American Journal of Community Psychology*, *26*(5), 775–802. doi:10.1023/A:1022162015815

Dvorsky, M. R., Girio-Herrera, E., & Owens, J. S. (2014). School-based screening for mental health in early childhood. In M. Weist, N. A. Lever, C. Bradshaw, & J. S. Owens (Eds.),

*Handbook of school mental health: Research, training, practice, and policy* (2nd ed., pp. 297–310). New York: Springer US. doi:10.1007/978-1-4614-7624-5_22

Edelbrock, C., Costello, A. J., Dulcan, M. K., Kalas, R., & Conover, N. C. (1985). Age differences in the reliability of the psychiatric interview of the child. *Child Development*, *56*(1), 265–275.

Edelsohn, G., Ialongo, N., Werthamer-Larsson, L., Crockett, L., & Kellam, S. (1992). Self-reported depressive symptoms in first-grade children: developmentally transient phenomena? *Journal of the American Academy of Child and Adolescent Psychiatry*, *31*(2), 282–290. doi:10.1017/S095457940000451X

Eklund, K., Renshaw, T. L., Dowdy, E., Jimerson, S. R., Hart, S. R., Jones, C. N., & Earhart, J. (2009). Early identification of behavioral and emotional problems in youth: Universal screening versus teacher-referral identification. *California School Psychologist*, *14*, 89–95.

Epkins, C. C. (1993). A preliminary comparison of teacher ratings and child self-report of depression, anxiety, and aggression in inpatient and elementary school samples. *Journal of Abnormal Child Psychology*, *21*(6), 649–661. doi:10.1007/BF00916448

Esch, P., Bocquet, V., Pull, C., Couffignal, S., Lehnert, T., Graas, M., … Ansseau, M. (2014). The downward spiral of mental disorders and educational attainment: a systematic review on early school leaving. *BMC Psychiatry*, *14*(1), 237. doi:10.1186/s12888-014-0237-4

Farmer, E. M. Z., Burns, B. J., Phillips, S. D., Angold, A., & Costello, E. J. (2003). Pathways into and through mental health services for children and adolescents. *Psychiatric Services*, *54*(1), 60–66. doi:10.1176/appi.ps.54.1.60

Farrington, D. P., & Loeber, R. (2000). Some benefits of dichotomization in psychiatric and criminological research. *Criminal Behaviour and Mental Health*, *10*(2), 100–122.

doi:10.1002/cbm.349

Feeney-Kettler, K. A., Kratochwill, T. R., Kaiser, A. P., Hemmeter, M. L., & Kettler, R. J. (2010). Screening young children's risk for mental health problems: A review of four measures. *Assessment for Effective Intervention*, *35*(4), 218–230. doi:10.1177/1534508410380557

Fergusson, D. M., & Horwood, L. J. (1992). Attention deficit and reading achievement. *Journal of Child Psychology and Psychiatry*, *33*(2), 375–85.

Flanagan, K. S., Bierman, K. L., & Kam, C. (2003). Identifying at-risk children at school entry: the usefulness of multibehavioral problem profiles. *Journal of Clinical Child and Adolescent Psychology*, *32*(3), 396–407. doi:10.1207/S15374424JCCP3203_08

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading*, *5*(3), 239–256. doi:10.1207/S1532799XSSR0503_3

Gadow, K. D., Kaat, A. J., & Lecavalier, L. (2013). Relation of symptom-induced impairment with other illness parameters in clinic-referred youth. *Journal of Child Psychology and Psychiatry*, *54*(11), 1198–1207. doi:10.1111/jcpp.12077

Gardner, W., Mulvey, E. P., & Shaw, E. C. (1995). Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. *Psychological Bulletin*, *118*(3), 392–404. doi:10.1037/0033-2909.118.3.392

Geiser, C. (2013). *Data analysis with Mplus*. New York: Guilford.

Girio-Herrera, E., Dvorsky, M. R., & Owens, J. S. (2015). Mental health screening in kindergarten youth: A multistudy examination of the concurrent and diagnostic validity of the Impairment Rating Scale. *Psychological Assessment*, *27*(1), 215–227.

doi:10.1037/a0037787

Gizer, I. R., Waldman, I. D., Abramowitz, A., Barr, C. L., Feng, Y., Wigg, K. G., … Rowe, D. C. (2008). Relations between multi-informant assessments of ADHD symptoms, DAT1, and DRD4. *Journal of Abnormal Psychology*, *117*(4), 869–880. doi:10.1037/a0013297

Glover, T. A., & Albers, C. A. (2007). Considerations for evaluating universal screening assessments. *Journal of School Psychology*, *45*, 117–135. doi:10.1016/j.jsp.2006.05.005

Good, R. H., & Kaminski, R. A. (2002). *Dynamic Indicators of Basic Early Literacy Skills (6th Edition)*. Eugene, OR. Retrieved from www.dibels.uoregon.edu

Goodman, A., Lamping, D. L., & Ploubidis, G. B. (2010). When to use broader internalising and externalising subscales instead of the hypothesised five subscales on the strengths and difficulties questionnaire (SDQ): Data from british parents, teachers and children. *Journal of Abnormal Child Psychology*, *38*, 1179–1191. doi:10.1007/s10802-010-9434-x

Goodman, R. (1994). A modified version of the Rutter parent questionnaire including extra items on children's strengths: A research note. *Journal of Child Psychology and Psychiatry*, *35*(8), 1483–1494. doi:10.1111/j.1469-7610.1994.tb01289.x

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, *38*(11), 581–586. doi:10.1097/00004583-200111000-00015

Goodman, R. (1999). The extended version of the Strengths and Difficulties Questionnaire as a guide to child psychiatric caseness and consequent burden. *Journal of Child Psychology and Psychiatry*, *40*(5), 791–799. doi:10.1111/1469-7610.00494

Goodman, R. (2001). Psychometric properties of the Strengths and Difficulties Questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry*, *40*(511), 1337–1345.

Goodman, R., Ford, T., Corbin, T., & Meltzer, H. (2004). Using the Strengths and Difficulties Questionnaire (SDQ) multi-informant algorithm to screen looked-after children for psychiatric disorders. *European Child and Adolescent Psychiatry, Supplement*, *13*(2004), 25–31. doi:10.1007/s00787-004-2005-3

Goodman, R., Ford, T., Simmons, H., Gatward, R., & Meltzer, H. (2000). Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *British Journal of Psychiatry*, *177*, 534–539. doi:10.1192/bjp.177.6.534

Goodman, R., & Scott, S. (1999). Comparing the Strengths and Difficulties Questionnaire and the Child Behavior Checklist: Is small beautiful? *Journal of Abnormal Child Psychology*, *27*(1), 17–24. doi:10.1023/A:1022658222914

Gordon, M., Antshel, K., Faraone, S., Barkley, R., Lewandowski, L., Hudziak, J. J., … Cunningham, C. (2006). Symptoms versus impairment: The case for respecting DSM-IV's Criterion D. *Journal of Attention Disorders*, *9*(3), 465–475. doi:10.1521/adhd.2005.13.4.1

Graham, J. M. (2006). Congeneric and (essentially) tau-equivalent estimates of score reliability. *Educational and Psychological MeasurementMeasurement*, *66*(6), 930–944. doi:10.1177/0013164406288165

Greenham, S. L. (1999). Learning disabilities and psychosocial adjustment: A critical review. *Child Neuropsychology*, *5*(3), 171–196. doi:10.1076/chin.5.3.171.7335

Grills-Taquechel, A. E., Fletcher, J. M., Vaughn, S. R., & Stuebing, K. K. (2012). Anxiety and reading difficulties in early elementary school: evidence for unidirectional- or bi-directional relations? *Child Psychiatry and Human Development*, *43*(1), 35–47. doi:10.1007/s10578-011-0246-1

Haapasalo, J., Tremblay, R. E., Boulerice, B., & Vitaro, F. (2000). Relative advantages of

person- and variable-based approaches for predicting problem behaviors from kindergarten assessments. *Journal of Quantitative Criminology*, *16*(2), 145–168.

Hagan-Burke, S., Kwok, O. -m., Zou, Y., Johnson, C., Simmons, D., & Coyne, M. D. (2011). An examination of problem behaviors and reading outcomes in kindergarten students. *The Journal of Special Education*, *45*(3), 131–148. doi:10.1177/0022466909359425

Halldorsdottir, T., & Ollendick, T. H. (2014). Comorbid ADHD: Implications for the treatment of anxiety disorders in children and adolescents. *Cognitive and Behavioral Practice*, *21*(3), 310–322. doi:10.1016/j.cbpra.2013.08.003

Hay, D. F., Pawlby, S., Sharp, D., Schmücker, G., Mills, A., Allen, H., & Kumar, R. (1999). Parents' judgements about young children's problems: why mothers and fathers might disagree yet still predict later outcomes. *Journal of Child Psychology and Psychiatry*, *40*(8), 1249–1258. doi:10.1111/1469-7610.00541

Headley, C., & Campbell, M. A. (2011). Teachers' recognition and referral of anxiety disorders in primary school children. *Australian Journal of Educational & Developmental Psychology*, *11*, 78–90.

Herman, K. C., Bi, Y., Borden, L. A., & Reinke, W. M. (2012). Latent classes of psychiatric symptoms among Chinese children living in poverty. *Journal of Child and Family Studies*, *21*(3), 391–402. doi:10.1007/s10826-011-9490-z

Herman, K. C., Lambert, S. F., Ialongo, N. S., & Ostrander, R. (2007). Academic pathways between attention problems and depressive symptoms among urban African American children. *Journal of Abnormal Child Psychology*, *35*(2), 265–74. doi:10.1007/s10802-006-9083-2

Herman, K. C., & Ostrander, R. (2007). The effects of attention problems on depression:

Developmental, academic, and cognitive pathways. *School Psychology Quarterly*, *22*(4), 483–510. doi:10.1037/1045-3830.22.4.483

Herman, K. C., Reinke, W. M., Parkin, J., Traylor, K. B., & Agarwal, G. (2009). Childhood depression: Rethinking the role of the school. *Psychology in the Schools*, *46*(5), 433–446. doi:10.1002/pits

Herrman, H. (2014). Early intervention as a priority for world psychiatry. *Early Intervention in Psychiatry*, *8*(4), 305–306. doi:10.1111/eip.12198

Hill, L. G., Coie, J. D., Lochman, J. E., & Greenberg, M. T. (2004). Effectiveness of early screening for externalizing problems: Issues of screening accuracy and utility. *Journal of Consulting and Clinical Psychology*, *72*(5), 809–820. doi:10.1037/0022-006X.72.5.809

Hong, S.-B., Dwyer, D., Kim, J.-W., Park, E.-J., Shin, M.-S., Kim, B.-N., … Cho, S.-C. (2014). Subthreshold attention-deficit/hyperactivity disorder is associated with functional impairments across domains: a comprehensive analysis in a large-scale community study. *European Child & Adolescent Psychiatry*, *23*(8), 627–636. doi:10.1007/s00787-013-0501-z

Hoza, B., Mrug, S., Gerdes, A. C., Hinshaw, S. P., Bukowski, W. M., Gold, J. a, … Arnold, L. E. (2005). What aspects of peer relationships are impaired in children with attention-deficit/hyperactivity disorder? *Journal of Consulting and Clinical Psychology*, *73*(3), 411–23. doi:10.1037/0022-006X.73.3.411

Huculak, S., & McLennan, J. D. (2014). Using teacher ratings to assess the association between mental health symptoms and impairment in children. *School Mental Health*, *6*, 15–26. doi:10.1007/s12310-013-9107-3

Hughes, J. N., Cavell, T. a., & Willson, V. (2001). Further support for the developmental significance of the quality of the teacher-student relationship. *Journal of School*

*Psychology*, *39*(4), 289–301. doi:10.1016/S0022-4405(01)00074-7

Hunsley, J. (2003). Introduction to the special section on incremental validity and utility in clinical assessment. *Psychological Assessment*, *15*(4), 443–445. doi:10.1037/1040-3590.15.4.443

Hunsley, J., & Mash, E. J. (2007). Evidence-based assessment. *Annual Review of Clinical Psychology*, *3*, 29–51. doi:10.1146/annurev.clinpsy.3.022806.091419

Hunsley, J., & Meyer, G. J. (2003). The incremental validity of psychological testing and assessment: conceptual, methodological, and statistical issues. *Psychological Assessment*, *15*(4), 446–455. doi:10.1037/1040-3590.15.4.446

Hymel, S., Rubin, K. H., Rowden, L., & Lemare, L. (1990). Children's peer relationships: Longitudinal prediction of internalizing and externalizing problems from middle to late childhood. *Child Development*, *61*(6), 2004–2021.

Ialongo, N., Edelsohn, G., Werthamer-Larsson, L., Crockett, L., & Kellam, S. (1995). The significance of self-reported anxious symptoms in first grade children: prediction to anxious symptoms and adaptive functioning in fifth grade. *Journal of Child Psychology and Psychiatry*, *36*(3), 427–437. doi:10.1111/j.1469-7610.1995.tb01300.x

Jarrett, M. A., & Ollendick, T. H. (2008). A conceptual review of the comorbidity of attention-deficit/hyperactivity disorder and anxiety: Implications for future research and practice. *Clinical Psychology Review*, *28*(7), 1266–80. doi:10.1016/j.cpr.2008.05.004

Jenkins, L. N., Demaray, M. K., Wren, N. S., Secord, S. M., Lyell, K. M., Magers, A. M., … Tennant, J. (2014). A Critical Review of Five Commonly Used Social-Emotional and Behavioral Screeners for Elementary or Secondary Schools. *Contemporary School Psychology*, *18*(4), 241–254. doi:10.1007/s40688-014-0026-6

Jenson, W. R., Harward, S., & Bowen, J. M. (2011). Externalizing disorders in children and adolescents: Behavioral excess and behavioral deficits. In M. A. Bray & T. A. Kehle (Eds.), *The Oxford handbook of school psychology* (pp. 379–410). New York: Oxford University Press.

Johnson, S., Hollis, C., Marlow, N., Simms, V., & Wolke, D. (2014). Screening for childhood mental health disorders using the Strengths and Difficulties Questionnaire: The validity of multi-informant reports. *Developmental Medicine and Child Neurology*, *56*(5), 453–459. doi:10.1111/dmcn.12360

Johnston, C., & Murray, C. (2003). Incremental validity in the psychological assessment of children and adolescents. *Psychological Assessment*, *15*(4), 496–507. doi:10.1037/1040-3590.15.4.496

Jones, D., Dodge, K. A., Foster, E. M., & Nix, R. (2002). Early identification of children at risk for costly mental health service use. *Prevention Science*, *3*(4), 247–256.

Kalberg, J. R., Lane, K. L., Driscoll, S., & Wehby, J. (2011). Systematic screening for emotional and behavioral disorders at the high school level: a formidable and necessary task. *Remedial and Special Education*, *32*(6), 506–520. doi:10.1177/0741932510362508

Kataoka, S. H., Zhang, L., & Wells, K. B. (2002). Unmet need for mental health care among U.S. children: variation by ethnicity and insurance status. *The American Journal of Psychiatry*, *159*(9), 1548–1555. doi:10.1176/appi.ajp.159.9.1548

Keiley, M. K., Bates, J. E., Dodge, K. A., & Pettit, G. S. (2000). A cross-domain growth analysis: Externalizing and internalizing behaviors during 8 years of childhood. *Journal of Abnormal Child Psychology*, *28*(2), 161–179.

King, K. R., Lembke, E. S., & Reinke, W. M. (2015). Using latent class analysis to identify

academic and behavioral risk status in elementary students. *School Psychology Quarterly*, Advance online publication. doi:http://dx.doi.org/10.1037/spq0000111

Kirby, M. J. L., & Keon, W. J. (2006). *Out of the Shadows at Last: Transforming Mental Health, Mental Illness and Addiction Services in Canada*. Retrieved from http://www.parl.gc.ca

Klasen, H., Woerner, W., Wolke, D., Meyer, R., Overmeyer, S., Kaschnitz, W., … Goodman, R. (2000). Comparing the German versions of the Strengths and Difficulties Questionnaire (SDQ-Deu) and the Child Behavior Checklist. *European Child and Adolescent Psychiatry*, *9*, 271–276. doi:10.1007/s007870070030

Kline, R. B. (2013). Exploratory and confirmatory factor analysis. In Y. Petscher & C. Schatschneider (Eds.), *Applied quantitative analysis in the social sciences* (pp. 171–207). New York: Routledge.

Kóbor, A., Takács, Á., & Urbán, R. (2013). The bifactor model of the strengths and difficulties questionnaire. *European Journal of Psychological Assessment*, *29*(4), 299–307. doi:10.1027/1015-5759/a000160

Koskelainen, M., Sourander, a., & Vauras, M. (2001). Self-reported strengths and difficulties in a community sample of Finnish adolescents. *European Child and Adolescent Psychiatry*, *10*, 180–185. doi:10.1007/s007870170024

Kovacs, M. (2011). *Children's depression inventory 2 (CDI-2)*. North Tonawanda, NY: Multi-Health Systems.

Kovacs, M., & Devlin, B. (1998). Internalizing disorders in childhood. *Journal of Child Psychology and Psychiatry*, *39*(1), 47–63. doi:10.1111/1469-7610.00303

Kratochwill, T. R., Albers, C. a., & Steele Shernoff, E. (2004). School-based interventions. *Child and Adolescent Psychiatric Clinics of North America*, *13*(4), 885–903.

doi:10.1016/j.chc.2004.05.003

Kremelberg, D. (2011). Logistic, ordered, multinomial, negative binomial, and poisson regression. In *Practical statistics: A quick and easy guide to IBM® SPSS® statistics, STATA, and other statistical software* (pp. 235–287). Thousand Oaks, CA: SAGE Publications. doi:10.4135/9781483385655.n6

Kuo, E., Vander Stoep, A., McCauley, E., & Kernic, M. a. (2009). Cost-effectiveness of a school-based emotional health screening program. *The Journal of School Health*, *79*(6), 277–285. doi:10.1111/j.1746-1561.2009.00410.x

Lane, K. L., Kalberg, J. R., Menzies, H., Bruhn, A., Eisner, S., & Crnobori, M. (2011). Using systematic screening data to assess risk and identify students for targeted supports: illustrations across the K-12 continuum. *Remedial and Special Education*, *32*(1), 39–54. doi:10.1177/0741932510361263

Lane, K. L., Menzies, H. M., Ennis, R. P., & Bezdek, J. (2013). School-wide systems to promote positive behaviors and facilitate instruction. *Journal of Curriculum and Instruction*, *7*(1), 6–31. doi:10.3776/joci.2013.v7n1p6-31

Lane, K. L., Menzies, H. M., Oakes, W. P., & Kalberg, J. R. (2012). *Systematic Screenings of Behavior to Support Instruction*. New York: Guilford.

Lane, K. L., Oakes, W., & Menzies, H. (2010). Systematic screenings to prevent the development of learning and behavior problems: Considerations for practitioners, researchers, and policy makers. *Journal of Disability Policy Studies*, *21*(3), 160–172. doi:10.1177/1044207310379123

Law, M., Hanna, S., Anaby, D., Kertoy, M., King, G., & Xu, L. (2014). Health-related quality of life of children with physical disabilities: A longitudinal study. *BMC Pediatrics*, *14*(1), 26.

doi:10.1186/1471-2431-14-26

Leech, S. L., Larkby, C. A., Day, R., & Day, N. L. (2006). Predictors and correlates of high levels of depression and anxiety symptoms among children at age 10. *Journal of the American Academy of Child and Adolescent Psychiatry*, *45*(2), 223–30. doi:10.1097/01.chi.0000184930.18552.4d

Levitt, J. M., Saka, N., Romanelli, L. H., & Hoagwood, K. (2007). Early identification of mental health problems in schools: the status of instrumentation. *Journal of School Psychology*, *45*(2), 163–191. doi:10.1016/j.jsp.2006.11.005

Lewin-Bizan, S., Bowers, E. P., & Lerner, R. M. (2010). One good thing leads to another: cascades of positive youth development among American adolescents. *Development and Psychopathology*, *22*(4), 759–70. doi:10.1017/S0954579410000441

Lilienfeld, S. O. (2003). Comorbidity between and within childhood externalizing and internalizing disorders: Reflections and directions. *Journal of Abnormal Child Psychology*, *31*(3), 285–291. doi:10.1023/A:1023229529866

Liu, X., Zhou, X., & Lackaff, J. (2013). Incremental validity in the clinical assessment of early childhood development. *Journal of Psychoeducational Assessment*, *31*(5), 506–515. doi:10.1177/0734282912473457

Loades, M. E., & Mastroyannopoulou, K. (2010). Teachers' recognition of children's mental health problems. *Child and Adolescent Mental Health*, *15*(3), 150–156. doi:10.1111/j.1475-3588.2009.00551.x

Lochman, J. E., & The Conduct Problems Prevention Research Group. (1995). Screening of child behavior problems for prevention programs at school entry. *Journal of Consulting and Clinical Psychology*, *63*(4), 549–559. doi:10.1037//0022-006X.63.4.549

Lopes, J., Cruz, C., & Rutherford, R. B. (2002). The relationship of peer perceptions to student achievement and teacher ratings of 5th and 6th grade students. *Education & Treatment of Children*, *25*(4), 476–495.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, *7*(1), 19–40. doi:10.1037/1082-989X.7.1.19

Magnusson, D. (2003). The person approach: Concepts, measurement models, and research strategy. *New Directions for Child and Adolescent Development*, (101), 3–23. doi:10.1002/cd.79

March, J. S. (2013). *Multidimensional anxiety scale for children (MASC-2)* (2nd Editio).

Margalit, M., & Zak, I. (1984). Anxiety and self-concept of learning disabled children. *Journal of Learning Disabilities*, *17*(9), 537–539. doi:10.1177/002221948401700906

Masten, A. S., & Cicchetti, D. (2010). Developmental cascades. *Development and Psychopathology*, *22*(3), 491–5. doi:10.1017/S0954579410000222

Masten, A. S., Roisman, G. I., Long, J. D., Burt, K. B., Obradovic, J., Riley, J. R., … Tellegen, A. (2005). Developmental cascades: linking academic achievement and externalizing and internalizing symptoms over 20 years. *Developmental Psychology*, *41*(5), 733–46. doi:10.1037/0012-1649.41.5.733

Maughan, B., & Carroll, J. (2006). Literacy and mental disorders. *Current Opinion in Psychiatry*, *19*(4), 350–4. doi:10.1097/01.yco.0000228752.79990.41

Maughan, B., Rowe, R., Loeber, R., & Stouthamer-Loeber, M. (2003). Reading problems and depressed mood. *Journal of Abnormal Child Psychology*, *31*(2), 219–229.

McCrory, C., & Layte, R. (2012). Testing competing models of the Strengths and Difficulties

Questionnaire's (SDQ's) factor structure for the parent-informant instrument. *Personality and Individual Differences*, *52*(8), 882–887. doi:10.1016/j.paid.2012.02.011

McCutcheon, A. L. (1987). *Latent Class Analysis*. Beverly Hills: SAGE Publications. doi:10.1016/j.jhsa.2013.01.024

McGee, R., Partridge, F., Williams, S., & Silva, P. A. (1991). A twelve-year follow-up of preschool hyperactive children. *Journal of the American Academy of Child and Adolescent Psychiatry*, *30*(2), 224–32. doi:10.1097/00004583-199103000-00010

McGrath, L. M., Pennington, B. F., Shanahan, M. A., Santerre-Lemmon, L. E., Barnard, H. D., Willcutt, E. G., … Olson, R. K. (2011). A multiple deficit model of reading disability and attention-deficit/hyperactivity disorder: searching for shared cognitive deficits. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *52*(5), 547–57. doi:10.1111/j.1469-7610.2010.02346.x

McLennan, J. D., Reckord, M., & Clarke, M. (2008). A mental health outreach program for elementary schools. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, *17*(3), 122–30.

McMichael, P. (1980). Reading difficulties, behavior, and social status. *Journal of Educational and Developmental Psychology*, *72*(1), 76–86.

Mental Health Commission of Canada. (2012). *Changing Directions Changing Lives: The Mental Health Strategy for Canada*. Retrieved from http://www.strategy.mentalhealthcommission.ca

Merikangas, K. R., He, J. P., Brody, D., Fisher, P. W., Bourdon, K. H., & Koretz, D. S. (2010). Prevalence and treatment of mental disorders among US children in the 2001-2004 NHANES. *Pediatrics*, *125*(1), 75–81. doi:10.1542/peds.2008-2598

Mieloo, C., Raat, H., van Oort, F., Bevaart, F., Vogel, I., Donker, M., & Jansen, W. (2012). Validity and reliability of the Strengths and Difficulties Questionnaire in 5-6 year olds: Differences by gender or by parental education? *PLoS ONE*, *7*(5). doi:10.1371/journal.pone.0036805

Mikami, A. Y., & Normand, S. (2015). The importance of social contextual factors in peer relationships of children with ADHD. *Current Developmental Disorders Reports*, *2*(1), 30–37. doi:10.1007/s40474-014-0036-0

Ministry of Health and Long Term Care. (2011). *Open Minds, Healthy Minds: Ontario's Comprehensive Mental Health and Addictions Strategy*. *Ministry of Health and Long Term Care*. Retrieved from http://www.health.gov.on.ca

Morgan, P. L., Farkas, G., Tufis, P. A., & Sperling, R. A. (2008). Are reading and behavior problems risk factors for each other? *Journal of Learning Disabilities*, *41*(5), 417–436. doi:10.1177/0022219408321123

Morrow, M. T., Hubbard, J. A., McAuliffe, M. D., Rubin, R. M., & Dearing, K. F. (2006). Childhood aggression, depressive symptoms, and peer rejection: The mediational model revisited. *International Journal of Behavioral Development*, *30*(3), 240–248. doi:10.1177/0165025406066757

Murphy, J. M., Guzmán, J., McCarthy, A. E., Squicciarini, A. M., George, M., Canenguez, K. M., … Jellinek, M. S. (2014). Mental health predicts better academic outcomes: A longitudinal study of elementary school students in chile. *Child Psychiatry & Human Development*, *46*(2), 245–256. doi:10.1007/s10578-014-0464-4

Murray-Close, D., Hoza, B., Hinshaw, S. P., Arnold, L. E., Swanson, J., Jensen, P. S., … Wells, K. (2010). Developmental processes in peer problems of children with attention-

deficit/hyperactivity disorder in the Multimodal Treatment Study of Children With ADHD: developmental cascades and vicious cycles. *Development and Psychopathology*, *22*(4), 785–802. doi:10.1017/S0954579410000465

Muthén, B. O., & Muthén, L. K. (2000). Integrating person-centered and variable-centered analyses: growth mixture modeling with latent trajectory classes. *Alcoholism, Clinical and Experimental Research*, *24*(6), 882–91. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10888079

Muthén, L. K., & Muthén, B. O. (2012). *User's Guide. Seventh Edition.* Los Angeles, CA.

Niclasen, J., Skovgaard, A. M., Andersen, A. M. N., Sømhovd, M. J., & Obel, C. (2013). A confirmatory approach to examining the factor structure of the strengths and Difficulties Questionnaire (SDQ): A large scale cohort study. *Journal of Abnormal Child Psychology*, *41*(3), 355–365. doi:10.1007/s10802-012-9683-y

Niclasen, J., Teasdale, T. W., Andersen, A. M. N., Skovgaard, A. M., Elberling, H., & Obel, C. (2012). Psychometric properties of the Danish strength and difficulties questionnaire: The SDQ assessed for more than 70,000 raters in four different cohorts. *PLoS ONE*, *7*(2), 1–8. doi:10.1371/journal.pone.0032025

Normand, S., Flora, D. B., Toplak, M. E., & Tannock, R. (2012). Evidence for a general ADHD factor from a longitudinal general school population study. *Journal of Abnormal Child Psychology*, *40*(4), 555–567. doi:10.1007/s10802-011-9584-5

Nottelmann, E. D., & Jensen, P. S. (1995). Comorbidity of disorders in children and adolescents. In T. H. Ollendick & R. J. Prinz (Eds.), *Advances in cliinical child psychology* (pp. 109–155). Boston, MA: Springer US. doi:10.1007/978-1-4757-9044-3

Nylund, K. L. (2007). *Latent Transition Analysis: Modeling Extensions and an Application to*

*Peer Victimization*. University of California Los Angeles. Retrieved from

https://www.statmodel.com/download/nylunddis.pdf

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(4), 535–569. doi:10.1080/10705510701575396

Offord, D. R., Boyle, M. H., Racine, Y., Szatmari, P., Fleming, J. E., Sanford, M., & Lipman, E. L. (1996). Integrating assessment data from multiple informants. *Journal of the American Academy of Child and Adolescent Psychiatry*, *35*(8), 1078–1085. doi:10.1097/00004583-199608000-00019

Offord, D. R., Boyle, M. H., Szatmari, P., Rae-Grant, N. I., Links, P. S., Cadman, D. T., … Woodward, C. A. (1987). Ontario Child Health Study II. Six-month prevalence of disorder and rates of service utilization. *Archives of General Psychiatry*, *44*, 832–836.

Okado, Y., & Bierman, K. L. (2014). Differential risk for late adolescent conduct problems and mood dysregulation among children with early externalizing behavior problems. *Journal of Abnormal Child Psychology*. doi:10.1007/s10802-014-9931-4

Ollendick, T. H., Jarrett, M. A., Grills-Taquechel, A. E., Hovey, L. D., & Wolff, J. C. (2008). Comorbidity as a predictor and moderator of treatment outcome in youth with anxiety, affective, attention deficit/hyperactivity disorder, and oppositional/conduct disorders. *Clinical Psychology Review*, *28*(8), 1447–1471. doi:10.1016/j.cpr.2008.09.003

Ontario Child Health Study. (2014). Ontario Child Health Study. Retrieved June 17, 2015, from http://ontariochildhealthstudy.ca/

Ostrander, R., & Herman, K. C. (2006). Potential cognitive, parenting, and developmental

mediators of the relationship between ADHD and depression. *Journal of Consulting and Clinical Psychology*, *74*(1), 89–98. doi:10.1037/0022-006X.74.1.89

Ostrander, R., Herman, K., Sikorski, J., Mascendaro, P., & Lambert, S. (2008). Patterns of psychopathology in children with ADHD: A latent profile analysis. *Journal of Clinical Child and Adolescent Psychology*, *37*(4), 833–47. doi:10.1080/15374410802359668

Owens, J. S., Storer, J., Holdaway, A. S., Serrano, V. J., Watabe, Y., Himawan, L. K., … Andrews, N. (2015). Screening for social, emotional, and behavioral problems at kindergarten entry: Utility and incremental validity of parent report. *School Psychology Review*, *44*(1), 21–40. doi:10.17105/SPR44-1.21-40

Pearson. (2013). Aimsweb. Retrieved from http://www.aimsweb.com/wp-content/uploads/Pearson_aimsweb_web_spread.pdf

Pemberton, C. K., Neiderhiser, J. M., Leve, L. D., Natsuaki, M. N., Shaw, D. S., Reiss, D., & Ge, X. (2010). Influence of parental depressive symptoms on adopted toddler behaviors: an emerging developmental cascade of genetic and environmental effects. *Development and Psychopathology*, *22*(4), 803–18. doi:10.1017/S0954579410000477

Pennington, B. F. (2006). From single to multiple deficit models of developmental disorders. *Cognition*, *101*(2), 385–413. doi:10.1016/j.cognition.2006.04.008

Pennington, B. F., Willcutt, E. G., & Rhee, S. H. (2005). Analyzing comorbidity. In R. V Kail (Ed.), *Advances in child development and behavior* (pp. 263–304).

Pianta, R. C., & Stuhlman, M. W. (2004). Teacher-child relationships and children's success in the first years of school. *Child Development*, *33*(3), 444–458.

Pickles, A., Rowe, R., Simonoff, E., Foley, D., Rutter, M., & Silberg, J. (2001). Child psychiatric symptoms and psychosocial impairment: relationship and prognostic significance. *British*

*Journal of Psychiatry, 179*, 230–235. doi:10.1192/bjp.179.3.230

Poncy, B. C., Skinner, C. H., & Axtell, P. K. (2005). An investigation of the reliability and standard error of measurement of words read correctly per minute using curriculum-based measurement. *Journal of Psychoeducational Assessment, 23*, 326–338.

Power, T. J., Andrews, T. J., Eiraldi, R. B., Doherty, B. J., Ikeda, M. J., DuPaul, G. J., & Landau, S. (1998). Evaluating attention deficit hyperactivity disorder using multiple informants: The incremental utility of combining teacher with parent reports. *Psychological Assessment, 10*(3), 250–260. doi:10.1037/1040-3590.10.3.250

Rabiner, D., & Coie, J. D. (2000). Early attention problems and children's reading achievement: A longitudinal investigation. *Journal of the American Academy of Child & Adolescent Psychiatry, 39*(7), 859–867. doi:10.1097/00004583-200007000-00014

Rapee, R. M., Bőgels, S. M., van der Sluis, C. M., Craske, M. G., & Ollendick, T. (2012). Annual research review: Conceptualising functional impairment in children and adolescents. *Journal of Child Psychology and Psychiatry, 53*(5), 454–68. doi:10.1111/j.1469-7610.2011.02479.x

Raykov, T. (1997). Scale reliability, Cronbach's coefficient alpha, and violations of essential tau-equivalence with fixed congeneric components. *Multivariate Behavioral Research, 32*(4), 329–353. doi:10.1207/s15327906mbr3204_2

Reinke, W. M., Herman, K. C., Petras, H., & Ialongo, N. S. (2008). Empirically derived subtypes of child academic and behavior problems: co-occurrence and distal outcomes. *Journal of Abnormal Child Psychology, 36*(5), 759–70. doi:10.1007/s10802-007-9208-2

Reynolds, C. R., & Kamphaus, R. W. (2004). *Behavior Assessment System for Children (BASC-2)* (2nd Ed.). Bloomington, MN: Pearson.

Reynolds, W. M., Anderson, G., & Bartell, N. (1985). Measuring depression in children: a multimethod assessment investigation. *Journal of Abnormal Child Psychology*, *13*(4), 513–526. doi:10.1007/BF00923138

Rijlaarsdam, J., Stevens, G. W., van der Ende, J., Hofman, A., Jaddoe, V. W., Verhulst, F. C., & Tiemeier, H. (2015). Prevalence of DSM-IV disorders in a population-based sample of 5- to 8-year-old children: the impact of impairment criteria. *European Child & Adolescent Psychiatry*. doi:10.1007/s00787-015-0684-6

Roberts, G. J., Solis, M., Ciullo, S., McKenna, J. W., & Vaughn, S. (2015). Reading interventions with behavioral and aocial skill outcomes: A synthesis of research. *Behavior Modification*, *39*(1), 8–42. doi:10.1177/0145445514561318

Romer, D., & McIntosh, M. (2005). The roles and perspectives of school mental health professionals in promoting adolescent mental health. In D. L. Evans, E. B. Foa, & R. E. Gur (Eds.), *Treating and preventing adolescent mental health disorders: What we know and what we don't know* (pp. 597–615). New York: Oxford University Press.

Roza, S. J., Hofstra, M. B., van der Ende, J., & Verhulst, F. C. (2003). Stable prediction of mood and anxiety disorders based on behavioral and emotional problems in childhood: A 14-Year follow-up during childhood, adolescence, and young adulthood. *American Journal of Psychiatry*, *160*(12), 2116–2121.

Russell, G., Ryder, D., Norwich, B., & Ford, T. (2015). Behavioural difficulties that co-occur with specific word reading difficulties: A UK population-based cohort study. *Dyslexia*, *21*, 123–141. doi:10.1002/dys.1496

Sayal, K., Owen, V., White, K., Merrell, C., Tymms, P., & Taylor, E. (2010). Impact of early school-based screening and intervention programs for ADHD on children's outcomes and

access to services: Follow-up of a school-based trial at age 10 years. *Archives of Pediatrics & Adolescent Medicine*, *164*(5), 462–469. doi:10.1001/archpediatrics.2010.40

School-Based Mental Health and Substance Abuse Consortium. (2013). *School-Based Mental Health in Canada: A Final Report*. Retrieved from http://www.mentalhealthcommission.ca

Schreiber, J. B., Stage, F. K., King, J., Nora, A., & Barlow, E. A. (2006). Modeling and confirmatory factor analysis results: A review. *Journal of Educational Research*, *99*(6), 323–337. doi:10.3200/JOER.99.6.323-338

Sechrest, L. (1963). Incremental validity: A recommendation. *Educational and Psychological Measurement*, *23*(1), 153–158.

Sellers, R., Maughan, B., Pickles, A., Thapar, A., & Collishaw, S. (2015). Trends in parent- and teacher-rated emotional, conduct and ADHD problems and their impact in prepubertal children in Great Britain: 1999-2008. *Journal of Child Psychology and Psychiatry*, *56*(1), 49–57. doi:10.1111/jcpp.12273

Severson, H. H., Walker, H. M., Hope-Doolittle, J., Kratochwill, T. R., & Gresham, F. M. (2007). Proactive, early screening to detect behaviorally at-risk students: Issues, approaches, emerging innovations, and professional practices. *Journal of School Psychology*, *45*(2), 193–223. doi:10.1016/j.jsp.2006.11.003

Sharp, C., Croudace, T. J., & Goodyer, I. M. (2005). The Strengths and Difficulties Questionnaire: predictive validity of parent and teacher ratings for help-seeking behaviour over one year. *Educational & Child Psychology*, *22*(3), 28–44.

Shemmassian, S. K., & Lee, S. S. (2012). Comparing four methods of integrating parent and teacher symptom ratings of attention-deficit/hyperactivity disorder (ADHD). *Journal of Psychopathology and Behavioral Assessment*, *34*(1), 1–10. doi:10.1007/s10862-011-9262-5

Simon, A. E., Pastor, P. N., Reuben, C. A., Huang, L. N., & Goldstrom, I. D. (2015). Use of mental health services by children ages six to 11 with emotional or behavioral difficulties. *Psychiatric Services*, appips201400342. doi:10.1176/appi.ps.201400342

Smith, J. P., & Smith, G. C. (2010). Long-term economic costs of psychological problems during childhood. *Social Science and Medicine*, *71*(1), 110–115. doi:10.1016/j.socscimed.2010.02.046

Sowerby, P., & Tripp, G. (2009). Evidence-based assessment of attention-deficit/hyperactivity disorder (ADHD). In J. L. Matson, F. Andrasik, & M. L. Matson (Eds.), *Assessing Childhood Psychopathology and Developmental Disabilities* (pp. 209–239). New York: Springer. doi:10.1007/978-0-387-09528-8

Stephan, S. H., Sugai, G., Lever, N., & Connors, E. (2015). Strategies for integrating mental health into schools via a multitiered system of support. *Child and Adolescent Psychiatric Clinics of North America*, 211–231. doi:10. 1 0 16/j.chc.2014.12.002

Stephan, S. H., Weist, M., Kataoka, S., Adelsheim, S., & Mills, C. (2007). Transformation of children's mental health services: The role of school mental health. *Psychiatric Services*, *58*(10), 1330–1338.

Sterba, S. K., & Bauer, D. J. (2010). Matching method with theory in person-oriented developmental psychopathology research. *Development and Psychopathology*, *22*(02), 239. doi:10.1017/S0954579410000015

Stone, L. L., Janssens, J. M. A. M., Vermulst, A. A., Van Der Maten, M., Engels, R. C. M. E., & Otten, R. (2015). The Strengths and Difficulties Questionnaire: Psychometric properties of the parent and teacher version in children aged 4-7. *BMC Psychology*, *3*(1), 4. doi:10.1186/s40359-015-0061-8

Stone, L. L., Otten, R., Engels, R. C., Vermulst, A., & Janssens, J. M. (2010). Psychometric properties of the parent and teacher versions of the Strengths and Difficulties Questionnaire for 4- to 12-Year-olds: A review. *Clinical Child and Family Psychology Review*, *13*, 254–274. doi:10.1007/s10567-010-0071-2

Stringaris, A., & Goodman, R. (2013). The value of measuring impact alongside symptoms in children and adolescents: A longitudinal assessment in a community sample. *Journal of Abnormal Child Psychology*, *41*(7), 1109–1120. doi:10.1007/s10802-013-9744-x

Sturaro, C., Van Lier, P. A. C., Cuijpers, P., & Koot, H. M. (2011). The role of peer relationships in the development of early school-age externalizing problems. *Child Development*, *82*(3), 758–765. doi:10.1111/j.1467-8624.2010.01532.x

Sullivan, H. S. (1953). *The interpersonal theory of psychiatry*. New York: Norton.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th Ed.). New York: Pearson.

Thorell, L. B., & Rydell, A. M. (2008). Behaviour problems and social competence deficits associated with symptoms of attention-deficit/hyperactivity disorder: effects of age and gender. *Child: Care, Health and Development*, *34*(5), 584–95. doi:10.1111/j.1365-2214.2008.00869.x

Tolan, P. H., & Henry, D. (1996). Patterns of psychopathology among urban poor children: comorbidity and aggression effects. *Journal of Consulting and Clinical Psychology*, *64*(5), 1094–9.

Tseng, W.-L., Kawabata, Y., Gau, S. S.-F., & Crick, N. R. (2014). Symptoms of Attention-Deficit/Hyperactivity Disorder and peer functioning: A transactional model of development. *Journal of Abnormal Child Psychology*, *42*(8), 1353–1365. doi:10.1007/s10802-014-9883-8

Tymms, P., & Merrell, C. (2006). The impact of screening and advice on inattentive, hyperactive and impulsive children. *European Journal of Special Needs Education*, *21*(3), 321–337. doi:10.1080/08856250600810856

Valdez, C. R., Lambert, S. F., & Ialongo, N. S. (2011). Identifying patterns of early risk for mental health and academic problems in adolescence: A longitudinal study of urban youth. *Child Psychiatry and Human Development*, *42*(5), 521–38. doi:10.1007/s10578-011-0230-9

van de Looij-Jansen, P. M., Goedhart, A. W., de Wilde, E. J., & Treffers, P. D. a. (2011). Confirmatory factor analysis and factorial invariance analysis of the adolescent self-report Strengths and Difficulties Questionnaire: How important are method effects and minor factors? *The British Journal of Clinical Psychology*, *50*, 127–144. doi:10.1348/014466510X498174

van Lier, P. A., & Koot, H. M. (2010). Developmental cascades of peer relations and symptoms of externalizing and internalizing problems from kindergarten to fourth-grade elementary school. *Development and Psychopathology*, *22*(3), 569–82. doi:10.1017/S0954579410000283

Vandeventer, M. C. (2008). *Universal Emotional and Behavioral Screening for Children and Adolescents: Are Multiple Gates and Informants Worth It?* University of Georgia.

Verhulst, F. C., Koot, H. M., & van der Ende, J. (1994). Differential predictive value of parents' and teachers' reports of children's problem behaviors: a longitudinal study. *Journal of Abnormal Child Psychology*, *22*(5), 531–546.

Waddell, C., Offord, D. R., Shepherd, C. A., Hua, J. M., & McEwan, K. (2002). Child psychiatric epidemiology and Canadian public policy-making: The State of the science and the art of the possible. *Canadian Journal of Psychiatry*, *47*(9), 825–832.

Wadsworth, M. E., Hudziak, J. J., Heath, A. C., & Achenbach, T. M. (2001). Latent class analysis of child behavior checklist anxiety/depression in children and adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry*, *40*(1), 106–14. doi:10.1097/00004583-200101000-00023

Walker, B. A. (2010). Effective schoolwide screening to identify students at risk for social and behavioral problems. *Intervention in School and Clinic*, *46*(2), 104–110. doi:10.1177/1053451210374989

Waschbusch, D. A. (2002). A meta-analytic examination of comorbid hyperactive-impulsive-attention problems and conduct problems. *Psychological Bulletin*, *128*(1), 118–150. doi:10.1037//0033-2909.128.1.118

Weare, K., & Nind, M. (2011). Mental health promotion and problem prevention in schools: What does the evidence say? *Health Promotion International*, *26*(SUPPL. 1). doi:10.1093/heapro/dar075

Weisz, J. R., Weiss, B., Alicke, M. D., & Klotz, M. L. (1987). Effectiveness of psychotherapy with children and adolescents: a meta-analysis for clinicians. *Journal of Consulting and Clinical Psychology*, *55*(4), 542–549. doi:10.1037/0022-006X.55.4.542

White, J., Connelly, G., Thompson, L., & Wilson, P. (2013). Assessing wellbeing at school entry using the Strengths and Difficulties Questionnaire: professional perspectives. *Educational Research*, *55*(1), 87–98. doi:10.1080/00131881.2013.767027

Wichstrøm, L., & Berg-Nielsen, T. S. (2014). Psychiatric disorders in preschoolers: The structure of DSM-IV symptoms and profiles of comorbidity. *European Child & Adolescent Psychiatry*, *23*(7), 551–62. doi:10.1007/s00787-013-0486-7

Wiener, J. (2008). Friendship and social adjustment of children with learning disabilities. In T.

H. Bryan, B. Y. Wong, & M. Donahue (Eds.), *The social dimensions of learning disabilities: Essays in honor of Tanis Bryan* (pp. 91–112). Mahwah, NJ.

Willcutt, E. G., Betjemann, R. S., Mcgrath, L. M., Chhabildas, N. A., Olson, R. K., DeFries, J. C., & Pennington, B. F. (2010). Etiology and neuropsychology of comorbidity between RD and ADHD: The case for multiple-deficit models. *Cortex*, *46*(10), 1345–1361. doi:10.1016/j.cortex.2010.06.009

Willcutt, E. G., & Pennington, B. F. (2000). Psychiatric comorbidity in children and adolescents with reading disability. *Journal of Child Psychology and Psychiatry*, *41*(8), 1039–1048.

Willcutt, E. G., Sonuga-Barke, E. J. S., Nigg, J. T., & Sergeant, J. A. (2008). Recent Developments in Neuropsychological Models of Childhood Psychiatric Disorders. In T. Banaschewski & L. A. Rohde (Eds.), *Biological Child Psychiatry: Recent Trends and Developments.* (Vol. 24, pp. 195–226). New York. doi:10.1159/000118526

Wille, N., Bettge, S., Wittchen, H.-U., & Ravens-Sieberer, U. (2008). How impaired are children and adolescents by mental health problems? Results of the BELLA study. *European Child & Adolescent Psychiatry*, *17 Suppl 1*(S1), 42–51. doi:10.1007/s00787-008-1005-0

Woerner, W., Fleitlich-Bilyk, B., Martinussen, R., Fletcher, J., Cucchiaro, G., Dalgalarrondo, P., … Tannock, R. (2004). The Strengths and Difficulties Questionnaire overseas: Evaluations and applications of the SDQ beyond Europe. *European Child and Adolescent Psychiatry, Supplement*, *13*(2004). doi:10.1007/s00787-004-2008-0

Wolff, J. C., & Ollendick, T. H. (2006). The comorbidity of conduct problems and depression in childhood and adolescence. *Clinical Child and Family Psychology Review*, *9*(3-4), 201–220. doi:10.1007/s10567-006-0011-3

World Health Organization. (1993). *The ICD-10 classification of mental and behavioral*

*disorders: diagnostic criteria for research*. Geneva, Switzerland: Author.

World Health Organization. (2013). *Mental Health Action Plan 2013-2020*. Retrieved from

http://apps.who.int/iris/bitstream/10665/89966/1/9789241506021_eng.pdf

World Health Organization. (2014). Mental health: strengthening our response. Retrieved from

http://www.who.int/mediacentre/factsheets/fs220/en/

Wu, J. Y., Hughes, J. N., & Kwok, O. M. (2010). Teacher-student relationship quality type in

elementary grades: Effects on trajectories for achievement and engagement. *Journal of*

*School Psychology*, *48*(5), 357–387. doi:10.1016/j.jsp.2010.06.004

Yates, B. T., & Taub, J. (2003). Assessing the costs, benefits, cost-effectiveness, and cost-benefit

of psychological assessment: we should, we can, and here's how. *Psychological*

*Assessment*, *15*(4), 478–495. doi:10.1037/1040-3590.15.4.478

Zahn-Waxler, C., Klimes-Dougan, B., & Slattery, M. J. (2000). Internalizing problems of

childhood and adolescence: prospects, pitfalls, and progress in understanding the

development of anxiety and depression. *Development and Psychopathology*, *12*(3), 443–

466. doi:10.1017/S0954579400003102