PERFORMANCE CALIBRATION IN CHILDREN WITH ATTENTION-DEFICIT/HYPERACTIVITY DISORDER (ADHD) AND A COMMUNITY SAMPLE OF TYPYCALLY DEVELOPING CHILDREN

ALEXANDRA BASILE

A DISSERTATION SUBMITTED TO

THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN PSYCHOLOGY

YORK UNIVERSITY

TORONTO, ONTARIO

AUGUST 2016

Abstract

This study investigated performance calibration in 81 8-12 year-old children [*M* = 119.89 months (or 9 years, 11months), *SD* = 14.92 months, 26 females], 39 with and 42 without ADHD, from the Greater Toronto Area. Performance calibration was evaluated across the domains of general knowledge, emotion recognition, and recognition of the valence of social situations using several indices. Parents of children in the clinical ADHD group rated their children's performance calibration skills significantly lower than parents in the typically developing (TD) group. While no group difference in performance calibration for general knowledge items were observed, the *easy-hard effect* was demonstrated, such that harder items elicited overconfidence and easier items elicited underconfidence across groups. The clinical ADHD group rated their effort expenditure significantly lower across most tasks although there were no differences in the perception of difficulty compared to the TD group. The clinical ADHD group was significantly overconfident on emotion recognition compared to the TD group, whereas the TD group was more accurate and underconfident across emotions. The clinical ADHD group was more confident overall when recognizing the valence of ambiguous social situations and more confident on the negative situations compared to the TD group. Thus, this study obtained performance calibration differences across facial emotion recognition and interpretation of social situations for children with ADHD compared to a TD group. The Bias Index, a measure of under/overconfidence, was recommended for future examinations of group differences in performance calibration. An investigation of the clinical utility of different performance calibration measures is presented and implications for interventions are discussed.

Acknowledgements

I would like to express my gratitude to the many people who have supported me throughout the completion of my dissertation. First, I would like to thank my supervisor and mentor, Dr. Maggie Toplak for her oversight, advise, support and faith in my ability to create such a novel project in our lab. Maggie, this six year journey through research and graduate school would not have been possible nor pleasurable without your guidance and for that I am thankful.

Second, I would like to thank my committee members, Drs. Brendan Andrade and John Eastwood, for their creativity, insightful comments and strong commitment to this project. I would like to thank Dr. Judy Weiner for taking on the role of external examiner, Dr. James Bebko for chairing the defense and loaning us a few participants, and Dr. Caroline Davis for participating as my internal/external member on my defense. The engagement and participation of all members made for an interesting, exciting, and memorable experience.

Third, this dissertation was a success due to the efforts set forth by my lab members and project co-creator Justine Ledochowski. I would like to thank Justine, Rachael Lyon, and Lee Propp for their help as co-data collectors, recruiters, and research assistants – I could not have asked for a better research team to complete the project at hand.

Finally, I would like to extend my appreciation to my husband and family for their continual love, encouragement, and support over these last six years. You have all been my cheerleaders and without your patience and reassurance this journey would not have been possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

Performance Calibration in Children with Attention-Deficit/Hyperactivity Disorder (ADHD) and a Community Sample of Typically Developing Children

Children tend to overestimate their competence on tasks (DeSoete & Roeyers, 2006; Schneider, Visé, Lockl, & Nelson, 2000). Estimating one's competence has been referred to as performance calibration; the metacognitive process of monitoring one's own thoughts or knowlege in relation to performance (Nelson, 1996). Performance calibration is an important skill because it has been negatively correlated with behavioural outcomes such as inhibition (Krueger et al., 2011). Children with Attention-Deficit/Hyperactivity Disorder (ADHD) have been reported to have a more limited awareness of their competence relative to typically developing (TD) children (Hoza et al., 2004). Studies demonstrate that children with ADHD experience difficulty with scholastic achievement, social interactions, athletics, and behaviour. Despite these challenges, these children retrospectively overestimate their performance in these domains compared to parent ratings, teacher ratings, and compared to TD children (Evangelista, Owens, Golden, & Pelham Jr., 2008; Hoza et al., 2004; Scholtens, Diamantopoulou, Tillma, & Rydell, 2012).

The purpose of the current study was to examine performance calibration in the domains of general knowledge, emotion recognition, and interpretation of social situations in children with ADHD and TD children aged 8 – 12 years old. The introduction will start by reviewing the literature on ADHD and commonly used measures of performance calibration with this population alongside the rationale for examining performance calibration across domains. Next, the theoretical underpinnings of the typical development of metacognition and performance calibration will be explained, followed by a review of the performance calibration literature including individual differences and task parameters that can impact performance calibration.

Several different terms have been used to describe performance calibration. Metacognition is a broad term that often describes the reflective process of one's thinking about their thinking. Performance calibration is a type of metacognition that infers thinking about one's performance on a task in relation to objective performance and can also encapsulate reports on self-perceptions. Under/overconfidence is a byproduct of performance calibration where one's estimated performance is better than (overconfidence) or worse than (underconfidence) their objective performance. The current study used the term metacognition to refer to broad reflective cognitive processes and the term performance calibration to refer to the comparison of self-perceived performance/confidence with a more objective measure of performance (for example, number of items correct).

Performance Calibration and ADHD

Attention-Deficit/Hyperactivity Disorder (ADHD) is a disorder of inattention and hyperactive-impulsive behaviour, which is typically first diagnosed in childhood (DSM-5, American Psychological Association, 2013) and often persists into adulthood (Polanczyk, de Lima, Horta, Bierderman, & Rohde, 2007; Willcutt, 2012). As a persistent neurodevelopmental disorder, ADHD effects between 5-7.1% of children and 2.5% of adults worldwide (Willcutt, 2012; Faraone et al., 2015). Some reports estimate 40% to 90% of children with ADHD will also have comorbid diagnoses of either Oppositional Defiant Disorder (ODD) or Conduct Disorder (CD) (Tannock, 1998) and therefore these symptoms were also measured in the current study. ADHD disproportionately affects males, and in childhood and adolescence, there is a 4:1 male-to-female ratio (Polanczyk et al., 2007; Willcutt, 2012). ADHD is often characterized by deficits in multiple cognitive domains underlying executive functions (EFs) such as difficulties with response inhibition or the ability to stop a prepotent response, working memory or the ability to mentally hold and manipulate information, and attention (Faraone et al., 2015; Seidman, 2006). Given these EF deficits, clinical assessments of ADHD typically rely upon parent and teacher reports of children's ADHD symptoms (Smith, Barkley, & Shapir, 2007). Children and adolescents with ADHD are not typically asked to self-report symptoms or difficulties associated with ADHD during assessment, which may be a reflection of the belief that they are impaired at self-monitoring. Lack of reliance on self-report of symptoms is unlike the diagnostic process for internalizing conditions, where children as young as age four have been relied upon to report symptoms that are the key factors in diagnostic practice (Luby, Belden, Sullivan, & Spitznagel, 2007; Smith et al., 2007). The value and use of self-report is not emphasized in practice guidelines for ADHD across Canada, Europe, the United States, and Australia (e.g., CADDRA, 2011; Pliszka, 2007; NHMRC, 2009; Taylor et al., 2004) which is based on the view that individuals with ADHD are impaired in self-monitoring.

While some adolescents with ADHD may appear self-aware in their academic self-appraisals, these self-appraisals may depend in part on EFs (Chan & Martinussen, 2015) which are important for self-awareness, self-reflection and perceptions of task difficulty and effort expenditure. Models of ADHD that describe cognitive processes core to the disorder include difficulties in self-awareness and reflection as important components of ADHD in addition to EF deficits (Barkley, 1996) including deficits in strategy use (Hervey, Epstein, & Curry, 2004). Most notably, self-perceptions of competence, a form of performance calibration, are rated significantly higher by individuals with ADHD than parent and teacher informants in scholastic, social, behavioural, and athletic domains (Evangelista et al., 2008; Hoza et al., 2004; Scholtens et al., 2012). Another area of self-reflection can involve perceptions of task difficulty and effort. One of the diagnostic criteria for ADHD is avoidance of effortful tasks and as such, may lead children with ADHD to perceive tasks to be more difficult and effortful than controls and therefore perceived effort and difficulty will also be used in the current study as an index of performance calibration.

In the ADHD literature, children with ADHD have demonstrated difficulty with self-monitoring, error detection, and self-report of symptoms (Balogh & Czobor, 2014; Smith et al., 2007). Performance calibration from the cognitive science literature has only been minimally examined in children and adults with ADHD with measures of under/overconfidence (DMC; Bruine de Bruin, Parker, & Fischhoff, 2007 and Y-DMC; Parker & Fischhoff, 2005). The Positive Illusory Bias is similar in nature to performance calibration paradigms, however, instead of using objective standards against which to compare performance (as in performance calibration), other's opinions or ratings are generally used as the reference point.

Positive Illusory Bias (PIB).

The *disparity* or *discrepancy* between self-report and informant report of competence is referred to as a Positive Illusory Bias (PIB; Hoza et al., 1993). Children with ADHD tend to rate their self-perceptions of competence in scholastic, social, athletic, and behaviour domains higher than their parents and teachers (Evangelista et al., 2008; Gerdes, Hoza, & Pelham, 2003; Hoza et al., 2004; Scholtens et al., 2012; Owens & Hoza, 2003). Discrepancy scores are larger and positive for individuals with ADHD, with parent/teacher ratings of children with ADHD being lower than children's self-ratings, which are consistently higher. Self-perceptions are inflated despite these children's documented difficulty with scholastic achievement, social interactions, athletics, and behaviour. Fefer, Off, and Dedrik (2015) found an association between students' overestimation of competence in social and academic domains and ADHD symptoms. When difference scores were used to categorize participants into underestimation, accurate, and overestimation groups, more than half of the study participants had biased self-perceptions and were either under/overconfident in their competence as compared to their teachers. However, ADHD symptoms were highest in students in the overestimation/overconfidence group.

For children with ADHD the PIB is manifested by elevated self-reports of competence compared to other informant reports of competence. In contrast to typical PIB studies, one study had children with ADHD complete a social/academic task and predict how they think they performed either before or after task completion. This study is very similar to performance calibration measures and the authors found that children with ADHD tend to give overconfident predictions relative to TD controls (Hoza et al., 2000; Hoza et al., 2001). Whether competencies are obtained based on informant reports or objective task performance, children with ADHD tend to be overconfident compared to typically developing children.

In one study, children with and without ADHD attended a 2-week summer day camp and prior to camp, their self- and parent/teacher perceptions of social and behavioural competence were assessed (Jia, Jiang, & Mikami, 2015). Discrepancy scores between child report and parent/teacher report for children with ADHD were uniquely predictive of oppositionality and peer preference at camp while only the informant ratings of competence were predictive of camp outcomes for the TD group. Another study had teenagers with ADHD participate in a simulated driving task and later rate their driving performance (Fabiano et al., 2015). Compared to other's ratings of their driving, the teenagers themselves exhibited a positive bias on their global ratings of driving ability. These positively biased ratings of global driving ability predicted greater rates of risky driving behaviour during the simulator exercise independent of disruptive behaviours. Thus, not only do children and adolescents with ADHD have positively biased self-perceptions, but these positively biased perceptions are also predictive of their behaviour.

Although PIB research has shed light on aspects of social-cognition that may be associated with children's social and behavioural functioning, the field is limited by methodological concerns; specifically, studies largely utilize informant comparisons. The sole use of informant comparisons is problematic for several reasons. In his commentary discussing 20 years of research on informant discrepancies, Achenbach (2011) summarizes two papers published in the 1980's that examined correlations between teacher and parent reports of childhood behaviour. The conclusion of one of these papers was that parents were poor informants because they did not agree with teachers, while the other paper concluded that teachers were poor informants because they did not agree with parents. The challenge with such measures is that the reference point for the behaviour being rated is ambiguous. Another way that performance calibration has been assessed in ADHD is through error processing.

Error processing.

The ability to detect and correct errors is an important aspect of controlling behaviour and has been found to be impaired in children with ADHD (O'Connell et al., 2009). Error processing has also been found to be a discrete component of executive control and this is supported by functional imaging that reveals dedicated neural networks for the detection and correction of performance errors (Fiehler, Ullsperger, & von Cramon, 2004; Garavan, Ross, Murphy, Roche, & Stein, 2002). Hypoactivation of these brain regions and a reduced tendency to correct one's performance post-error has been found in children with ADHD (Balogh & Czobor, 2014; O'Connell et al., 2009; Schachar et al., 2004; and van Meel, Heslenfeld, Oosterlaan, & Sergeant, 2007).

In one study, adults with and without ADHD completed a Go/No-Go response inhibition task that was used to assess error awareness (O'Connell et al., 2009). The ADHD group made significantly more errors and were less likely to consciously detect these errors compared to the control group. Additionally, the event-related potentials elicited by errors that typically indicate early performance monitoring as well as a later component reflective of conscious error processing was significantly attenuated in the ADHD group as shown by decreased activation of the anterior cingulate cortex. Similarly, post-error slowing has been used experimentally to assess self-monitoring in children with ADHD (Schachar et al., 2004; Balogh & Czobor, 2014). Post-error slowing describes the tendency to slow down after making an error on the previous trial of a cognitive task. Schachar and colleagues (2004) found that compared to controls, participants with ADHD slowed less after errors on a stop signal reaction time task. This result has also been replicated in children aged 8-12 years old with and without ADHD (van Meel, Heslenfeld, Oosterlaan, & Sergeant, 2007) indicating that children with ADHD may possess a specific deficit in monitoring ongoing behaviour. Meta-analyses of post-error slowing indicate that individuals with ADHD tend to show less post-error slowing compared to TD individuals (Balogh & Czobor, 2014) which indicates that they do not adequately self-monitor performance on tasks.

Measures of under/overconfidence.

Performance calibration has also been assessed by measures of under/overconfidence. Under/overconfidence has been included and assessed as a component of the Adult Decision-Making Competence measure (DMC; Bruine de Bruin, Parker, & Fischhoff, 2007) and the Youth Decision-Making Competence measure (Y-DMC; Parker & Fischhoff, 2005). Under/overconfidence is assessed by having participants answer general knowledge questions and then rate their confidence in the accuracy of their answers. Mäntylä and colleagues (2012) examined performance calibration with an under/overconfidence task in adults with ADHD and community controls. In this study, under/overconfidence did not significantly predict ADHD status although another DMC task (applying decision rules) did predict ADHD status. The relationship between ADHD and decision-making competence tasks is unclear and consistent with previous findings that ADHD status does not have a clear influence on metacognitive functions (Knouse, Paradise, & Dunlosky, 2006; Weller, Levin, Rose, & Bossard, 2012).

It is surprising that under/overconfidence does not predict ADHD status given that there is so much evidence to suggest that people with ADHD are poor self-monitors of performance (see PIB; Evangelista et al., 2008; Hoza et al., 2004; Scholtens et al., 2012). Unfortunately, the DMC overconfidence task (essentially performance calibration) has not been examined in children with ADHD. Under/overconfidence has been significantly associated with attention focusing and inhibitory control and therefore it would be expected that since children with ADHD have impairments with these EFs, they would also show impairment with under/overconfidence. Since there is only one study that assesses performance calibration through under/overconfidence in adults with ADHD in this objective concurrent way and there is no research that examines this in children, more research is needed to explain the relationship between ADHD and performance calibration in children. The current study utilized a performance calibration paradigm across general knowledge, affect recognition, and interpretation of social situations tasks to examine how children with and without ADHD calibrate performance to determine whether calibration differs between groups and domains.

Overall, experimentally and clinically, deficits in the calibration of behaviour and performance have been implicated in the ADHD literature with respect to the PIB (Evangelista et al., 2008; Hoza et al., 2004; Scholtens et al., 2012), error detection (Balogh & Czobor, 2014; O’Connell et al., 2009; Schachar et al., 2004), and only minimally with overconfidence paradigms (Mäntylä et al., 2012). Based on this review, it was expected that children with ADHD would have significantly poorer performance calibration compared to the TD group. Moreover, since the diagnostic criteria for ADHD include avoidance of effortful tasks, children with ADHD were expected to find tasks more difficult and effortful than controls. Specifically, performance calibration across the domains of general knowledge, emotion recognition, and interpretation of social situations were examined in this study.

Performance Calibration across Domains

The current study assessed performance calibration in three domains: general knowledge, emotion recognition, and interpretation of social situations.

General knowledge.

In the adult literature, measurement of under/overconfidence has been well-established (Bruine de Bruin, Parker, & Fischoff, 2007). Some research has examined under/overconfidence in preadolescent TD children using general knowledge questions (Weller et al., 2012). Weller and colleagues (2012) examined the use of the DMC measure in typically developing preadolescents. Children’s confidence scores indicated that their performance calibration was moderately accurate. Children with ADHD have been found to perform more poorly than typically developing children on tests general knowledge such as the Information subtest on the WISC-R (Loge, Staton, & Beatty, 1990) therefore it is reasonable to hypothesize that children with ADHD should perform worse on general knowledge tasks than a TD control group. The proposed study aimed to extend these finding by using a general knowledge task to compare performance calibration between TD children and a clinical sample of children with ADHD, since no research exists that compares performance calibration on this task between these children.

Emotion recognition.

Facial expressions are a nonverbal means to communicate and express one's own emotions as well as to recognize others' emotional states (Singh et al., 1998). Sensitivity to facial cues of emotional expression is imperative for children's social competence (Chronaki, Hadwin, Garner, Maurage, & Sonuga-Barke, 2015). Children who understand facial affect in social interactions have also been found to form positive interpersonal relationships (Denham et al., 2013). At as young as four months of age, infants are able to discriminate between happiness and anger (Barrera & Maurer, 1981) and can adjust their behaviour to the emotion depicted in a facial expression (Hertenstein & Campos, 2004). Affect recognition and facial emotional processing continues to develop from preschool to adolescence (Herba, Landau, Russell, Ecker, & Phillips, 2006) with sadness and anger being the least accurately recognized facial emotions relative to happiness, fear and disgust in 4-16 year-olds (Montirosso, Peverelli, Frigerio, Crespi & Borgatti, 2010). In TD samples, children's emotion recognition has been shown to improve with age and reach adult levels by age 11 (Chronaki et al., 2015). Compared to anger and happiness, sadness recognition is acquired later in development. Thus, results from developmental studies suggest that there are different developmental trajectories for different emotions.

There have been three different explanations and lines of research to explain why children with ADHD perform poorly on recognition of facial expressions: 1) A neurodevelopmental explanation that posits that the underlying problem is due to inattention and impulsivity (Draeger, Prior, & Sanson, 1986); 2) A social learning explanation which contends that since accuracy in ADHD differs by emotion, facial emotion recognition may be highly influenced by social experiences (McAlpine et al., 1992); and 3) A facial processing explanation which suggests that adolescents with ADHD process emotional facial stimuli differently than do adolescents without ADHD (Dan & Raz, 2015).

Research suggests that children with ADHD may be impaired in their ability to accurately recognize facial expressions of emotions due to their inattention and impulsivity (Singh et al., 1998). Difficulties with inattention and impulsivity in ADHD may lead children to make more errors in emotion recognition due to a failure to attend to facial cues or because of responses made based on incorrect social judgments. In order to interpret facial cues as emotions, different parts of the face must be attended to as different parts can provide different cues (McAlpine, Singh, Ellis, Kendall, & Hampton, 1992). When children with ADHD, children with conduct problems (CP; 2+ DSM-IV CD symptoms met), and TD controls were asked to interpret emotional cues of facial expressions using picture and voice recordings, children with CP and ADHD were significantly less accurate at interpreting emotions than the TD controls (Cadesky, Mota, & Schachar, 2000). Findings from this study provide some evidence that social problems in ADHD may be partly associated with deficits in the ability to attend to appropriate social cues. This may be different from the social problems of children with CP who tend to make biased judgments related to the misinterpretation of emotions as angry. These findings are corroborated by a study of young adults diagnosed with childhood ADHD who made fewer correct responses to facial expressions despite showing similar brain activation to controls without childhood ADHD (Schulz et al., 2014). Indeed, in visual tasks that require attention, hyperactive children may perform poorly as a result of deficits in their ability to attend to the appropriate social cues (Draeger, Prior, & Sanson, 1986).

In another study, Singh and colleagues (1998) read a story to children depicting an emotion and asked children aged 5-13 year-old with ADHD to identify which of six photos portraying the six basic emotions (sad, happy, surprised, frightened, angry, and disgusted) matched the story. These authors found that children with ADHD had deficits in their ability to accurately recognize and match facial expressions to stories, with an accuracy of 74% as compared to children from another study from the general population whose accuracy was 89%. The ADHD sample correctly identified 94% of the happiness emotions and 86% of sadness while the general population of children identified 100% and 88% respectively. Similarly, other studies indicate that boys with ADHD have significant impairment in facial emotion interpretation, specifically with depictions of fear and disgust (Boakes, Chapman, Houghton, & West, 2008). Others have instead purposed that the recognition of facial expression may be related to exposure to that emotion (McAlpine et al., 1992) and therefore children with externalizing disorders have more exposure to disapproval from peers and adults thus giving them more opportunities to learn negative facial expressions such as anger and disgust (Singh et al., 1998).

Dan and Raz (2015) investigated emotional face processing in adolescents aged 15-18 years old with and without ADHD. Participants were asked to rate the degree of positivity/negativity of four facial expressions and their ratings, rating variability, response time, and response time variability were analyzed. Compared to adolescents without ADHD, adolescents with ADHD were less able to discriminate between positive and negative emotional expressions and showed greater variability in their reaction times and in the degree of their positivity/negativity ratings of facial expressions. In line with these findings, Serrano and colleagues found that for some emotions, children with ADHD spent less time viewing relevant areas of a picture and took longer to detect emotions compared to children without ADHD (Serrano, Owens, & Hallowell, 2015). Specifically, Dan and Raz (2015) found that participants with ADHD perceived emotional stimuli as similar to neutral expressions, differing from participants without ADHD who rated emotions as extremely positive or negative. Moreover, participants with ADHD rated happy and neutral facial expressions as less positive than the control group suggesting that adolescents with ADHD process happy faces differently from typically developing adolescents. These results suggest a specific impairment in the processing of emotional face stimuli in adolescents with ADHD.

It has been demonstrated that children and adolescents with ADHD have impaired attention, different social experiences and learning, and different processing of emotional face stimuli which in turn may lead to impaired facial emotion recognition and this has been echoed in a systematic review of facial emotion recognition (Collin, Bindra, Raju, Gillberg, & Minnis, 2013). No studies to date have examined children's confidence in the accuracy of their emotion recognition. This is an important area of study as it may help to shed some light on how children think about and reflect on their emotion recognition. The current study built on the existing literature on deficits in emotion recognition in ADHD by examining the confidence ratings for recognition of different emotions on facial expressions in children with and without ADHD. Impairments in facial emotion interpretation are also a source of impaired social competence (Simon, Rosen, Grossman, & Pratowski, 1995). Since confidence for accuracy of emotion recognition is a new area of research, performance calibration across the five emotions presented in the Emotion Recognition Task was explored in this study.

Interpretation of social situations: Social cognition.

Middle childhood is a time when children enhance their socioemotional competence which can impact peer acceptance, self-esteem and various developmental outcomes (Gifford-Smith & Browness, 2003). Deficits in social information processing (SIP) are identified as a fundamental to maladaptive interpersonal functioning for many children with ADHD (Whalen & Henker, 1985). According to SIP theory, children process social information by 1) encoding external cues, 2) interpreting attributions regarding self and others, 3) clarifying their goals, 4) generating possible responses, 5) choosing a response, and 6) enacting the response (Crick & Dodge, 1994). SIP models have typically been applied to the interpretation of children's aggressive behaviour and more recently to interpret the social-cognitive abilities of children with ADHD (Andrade et al., 2012). The relevant aspects of the SIP theory for the current study are the encoding of external cues and interpretation of attributions in order to accurately interpret the valence of social situations, which has also been described as social cognition (Uekermann et al., 2010). Social cognition is further defined as the encoding, representation, and interpretation of social cues (Uekermann et al., 2010).

Andrade and colleagues (2012) asked children with and without ADHD, aged 6-12 years, to listen to vignettes that had ambiguous, positive, and negative outcomes and peer intentions in order to evaluate their SIP abilities. Results indicated that children with ADHD showed impaired detection and encoding of external cues, such that they attributed more negative and less positive intent to peers. Moreover, children with ADHD relied less on the situational outcomes of the vignettes and had less positive responses to the situations as compared to the control group. Furthermore, another study found that boys with ADHD, ODD or Conduct Disorder (CD), and ODD/CD + ADHD encoded fewer social cues and generated fewer responses to videotaped stimuli of social situations than did a TD control group (Matthys, Cuperus, & Engeland, 1999). These findings suggest that children with ADHD are impaired in detection and encoding of cues and respond incongruently to situations, possibly more negatively than do TD children.

As indicated earlier, facial emotion recognition is impaired in children with ADHD, which may lead to more errors in emotional recognition (Singh et al., 1998). Impairment in facial emotion recognition for children with ADHD may lead to other social challenges. Children with ADHD showed deficiencies in handling conflict and maintaining friendships (Grenell, Glass, & Katz, 1987), were rated lower in terms of social preference, were less well liked, more often rejected/less accepted, had fewer friendships, poor quality friendships, spend less time with friends out of school, and were more likely to be described as non-friends by other children with higher social status (Hoza et al., 2005; Marton, Wiener, Rogers, & Moore, 2015; Mikami & Normand, 2015). Indeed, impairments in detecting emotions from facial expression are associated with low social competence and low peer popularity (Edwards, 1984; Philippot & Feldman, 1990). Hinshaw and Erhardt (1992) said that “the interpersonal problems of children with ADHD may well constitute the most salient and debilitating aspects of their psychopathologic behaviour” (p. 539). Further research supports the assertion that children with ADHD exhibit interpersonal interactions that are conflictual and negative (Barkley, 2006; Green et al., 2001).

Moreover, the PIB literature has also suggested that children with ADHD have poor social interactions and rate themselves more socially competent than do other informant reports (Hoza 2007; Hoza et al., 2005). People with ADHD often have a range of social and interpersonal challenges that can result in social problems (Nijmeijer et al., 2008). Children with ADHD do not lack interest in engaging with other people, but rather, have difficulties adjusting their behaviour to people (Nijmeijer et al., 2008). This adjusting can be viewed as a means of performance calibration, in terms of calibrating one’s behaviour to the perceived expectations of others. It may be the case that children with ADHD start out with an initial deficit in recognizing facial emotions; however, they still must attempt to match their responses to social situations with incorrect or missing information, and this in turn may bolster miscalibration. Therefore, it is expected that children with ADHD should be impaired in the encoding of external cues and interpretation of the valence of social situations since these children also show impairment in their social cognition and interpersonal relationships. The current study built upon the existing literature for deficits in SIP in ADHD by examining the confidence ratings for interpretation of the valence of social situations in children with and without ADHD. Confidence ratings for accuracy of identifying ambiguous positive and ambiguous negative social situations were explored. Ambiguous social situations are intended to evoke a feeling of ambiguity which should in turn decrease confidence or lead to underconfidence and elicit personal judgments. In order to understand the typical development of metacognition and more specifically, performance calibration, a review of the literature is presented below.

Typical Development of Metacognition and Performance Calibration

Metacognition has been defined as “knowing about knowing” or “cognitions about cognitions” (Barell, 1992) and are second-order skills that include knowing one’s own and others’ knowing (Kuhn, 2014). Indeed, metacognition has been integrally linked with most definitions of critical thinking (Kuhn, 2014) and understanding how children make decisions can provide insight into the development of their rational thought (Weller et al., 2012). When individuals need to make a decision, their thoughts about their thinking allows them to select and apply a strategy, monitor performance, revise the strategy, and evaluate performance in various domains (Krueger et al., 2011). This last step of decision-making, evaluation of performance, is the metacognitive process of performance calibration; a person’s judgment of his or her performance and his or her actual performance (Keren, 1991). Accordingly, Kuhn (1989) proposed that knowledge acquisition is the result of the coordination of existing understandings with new evidence. Over the course of development, children gain increasing control over coordination of these processes. Since this development requires awareness, understanding, and control over one's cognition, it involves the attainment of meta-knowing (Kuhn, 2014).

Kuhn (2014) divides meta-knowing into three categories; 1) Metacognitive, 2) Metastrategic, and 3) Epistemological. Metacognitive knowing operates on the basis of declarative knowledge, or "knowing that" which benefits from executive management or rather what is known and how it is known, while metastrategic knowing involves the selection and monitoring of applied strategies. Lastly, epistemological knowing is a broader sense of understanding that is both general (i.e. How does anyone know?) and also personal (i.e. What do I know about my knowing?). Meta-cognitive knowing and metastrategic knowing are both involved in the developmental of performance calibration and are therefore the focus of this review.

Metacognitive knowing is thought to develop within the three to five year age range, when children are believed to acquire the insight that when people make assertions, they are reflections of someone's belief (Olson & Astington, 1993) which is a marker for the development of meta-knowing and essential for the later development of critical thinking (Kuhn, 2014). *Metastrategic knowing* has not yet been developed in five-year-old children and instead, is believed to develop at approximately middle childhood (Kuhn, 2014). For instance, during preschool years, children tend not to spontaneously apply strategies such as mnemonic devices or semantic pairing (i.e. tree goes with green) when they may be useful, such as when completing memory tasks (i.e. recalling word lists); while Grade 3 and 5 students are more strategic and self-aware of the problems, strategies, and solutions to help with recall and retrieval (Kreutzer, Leonard, & Flavell, 1975). Change or development in metastrategic knowing occurs in continuing shifts in the frequency and usage of strategies, the elimination of less successful strategies, and finally, strategy selection from an individual's learned set of strategies (Kuhn, 2014).

A major component of cognitive development is how thought becomes more aware of itself and is also under one's own control (Kuhn, 2014). In order to take control of one's life, taking control of one's own thinking is often an important step. Metacognitive, metastrategic, and epistemological knowing are central to the development of critical thinking as critical thinking involves reflecting on knowledge and how that knowledge is justified. One way in which children may be able reflect on their own knowledge is through self-monitoring, or self-perception of performance on various tasks in relation to objective performance on those tasks. This is termed, *performance calibration*. Some studies have found that performance on under/overconfidence tasks improve with age (Bruine de Bruin et al., 2007). Children display overconfidence in relation to estimates of their performance on mathematics tests and get better at evaluating their performance on math tasks with age (Desoete & Roeyers, 2006).

Performance Calibration Paradigms from the Cognitive Science Literature

A higher degree of fit between actual performance on a task and an individual's judgment of his or her performance indicates better performance calibration (Keren, 1991). The under/overconfidence task of the Pre-Adolescent Decision-Making Competence battery requires individuals to complete general knowledge items and then make confidence judgments after each item (Weller et al., 2012). The confidence judgments are compared to the correctness of the answer provided to determine whether confidence is in line with performance. The general findings of such studies with adults is that people are consistently overconfident, so much so, that they are even willing to wage money on their validity (Fischhoff, Slovic, and Lichtenstein, 1977). What, then, causes people to be so miscalibrated?

Klayman and Soll (1999) offer two alternate explanations for why people tend to be under/overconfident: (a) biased information processing and (b) effects of judgmental error. These two reasons for under/overconfidence are similar to the SIP deficit described above with regards to ADHD; children with ADHD have impaired attention, biased social learning, impaired facial processing and tend to make impaired/incongruent responses to situations. In the biased information processing model, when people are asked a question, they search their memory for relevant information and generate an answer. People then search their memory for evidence to support that answer. Mechanisms of associative memory retrieval of information lead people to remember information that is consistent with their first answer and also influence the interpretation of ambiguous evidence in the direction of the initial answer. People are unaware that this process is biased and therefore people perceive the information retrieved as more support for the initial answer (Klayman, 1995; Hoch, 1985).

The second explanation for under/overconfidence is the role of judgmental error (Klayman & Soll, 1999). Errors can be produced when evaluating the information available (Erev, Wallsten, & Budescu, 1994), when learning the predictive validity of different sources of information (Gigerenzer, Hoffrage, & Kleinbölting, 1991), and when mapping the subject feeling of confidence to a response scale (Erev et al., 1994). Both accuracy and confidence can be affected by random variation (Klayman & Soll, 1999). Since good-quality information has the chance of leading us in the wrong direction, the frequency of that occurring is also a product of chance. People's judgment of the quality of their information also includes error and can influence confidence. Moreover, there is an imperfect correlation between accuracy and confidence and therefore it is not surprising that low accuracy is associated with "not-so-low confidence" producing a pattern of miscalibration (Klayman & Soll, 1999). Confidence also varies by question, by individual, and by domain and consequently, these factors were explored in the current study.

Performance calibration and cognitive abilities.

A positive association between cognitive abilities and resistance to overconfidence has been reported (Stanovich & West, 2000; Wolfe & Grosch, 1990). For instance, general verbal ability has been found to predict performance calibration on readings tasks (Gillstrom & Ronnberg, 1995). Adolescents' performance on under/overconfidence tasks are positively associated with a general EF measure (Parker & Fischhoff, 2005).Weller and colleagues (2012) examined the relationship between under/overconfidence and cognitive abilities such as inhibitory control, the ability to stop a prepotent response. Greater inhibitory control was significantly associated with the accuracy of confidence ratings indicating that better EFs may be associated with better performance calibration. Dougherty and Sprenger (2006) found that working memory capacity was related to the metacognitive task of making probability judgments, while Souchay, Isingrini, Clarys, and Eustache (2004) found that relative accuracy of feelings of knowing were correlated with EF functions but that judgments of learning were not. The relationship between cognitive abilities including EFs and performance calibration is not yet clear. The current study aimed to clarify the roles of cognitive ability and EF in performance calibration among children. It was expected that children who have higher cognitive abilities (estimated intelligence and EFs) would display more accurate performance calibration than those with lower cognitive abilities.

Reliability of performance calibration measures.

Some researchers have proposed ways to assess performance calibration reliably (Thompson, 1999; Veenman, Van Hout-Walters, & Afflerback, 2006). There are many reasons why measurements of performance calibration may have low reliability including few judgments, time of assessment, and grain size. A score based on a low number of judgments may be one reason for low reliability and that reliability will be maximized as the number of judgments increases past four to six items and has a sufficient range of difficulty (Osterlind, 2006; Weaver, 1990). Delayed judgments of learning (i.e. delayed judgments made after the all items are completed, JOL) are more accurate than judgments made immediately after the item is complete because of cognitive factors that can be employed to reflect on aggregate performance (Thiede, Dunlosky, Griffin, & Wiley, 2005; Veenman et al., 2006). Moreover, Schneider and colleagues (2000) had young children complete metamemory judgments of learning and also found that overconfidence was larger for item-by-item JOLs than for aggregate JOL. Lastly, grain size refers to how many items are used to calibrate performance. Composite judgments that require judgments be made on a set of items instead of individual items can be problematic because reliability cannot be made based on a single rating and even when multiple data points (i.e. number of items answered correct) are used, one’s estimates of performance may be adjusted to match a sense of average performance (Dunlosky, Rawson, & Middleton, 2005; Schraw, 2008, 2010).

The current study addresses reliability challenges in the following ways. More than six judgments/items were used for each domain of performance calibration measured. Both immediate/concurrent judgments of performance (confidence ratings) and delayed/retrospective judgments (effort and perceived difficulty) were obtained from participants. Reliability and validity of the performance calibration measured employed were also evaluated. Finally, indices of performance calibration selected for the current study were chosen based on the construct being measured. Thus, the current study follows the recommendations for practice set out by Schraw (2008, 2010). The reliability of the indices used are described in the methods section.

Measurement of performance calibration in the current study.

The measurement of performance calibration in the current study was based on what Schraw (2010) classifies as a concurrent judgment and retroactive judgments about the accuracy of performance. Concurrent confidence judgments are made immediately on an item-by-item basis. Retrospective judgments are made after the completion of an entire task. Measuring performance calibration in this way provides multiple measures of the ability to monitor performance on various tasks. Moreover, two indices of absolute accuracy and two indices of relative accuracy were employed; absolute and relative accuracy (Schraw, 2010).

Absolute and relative accuracy.

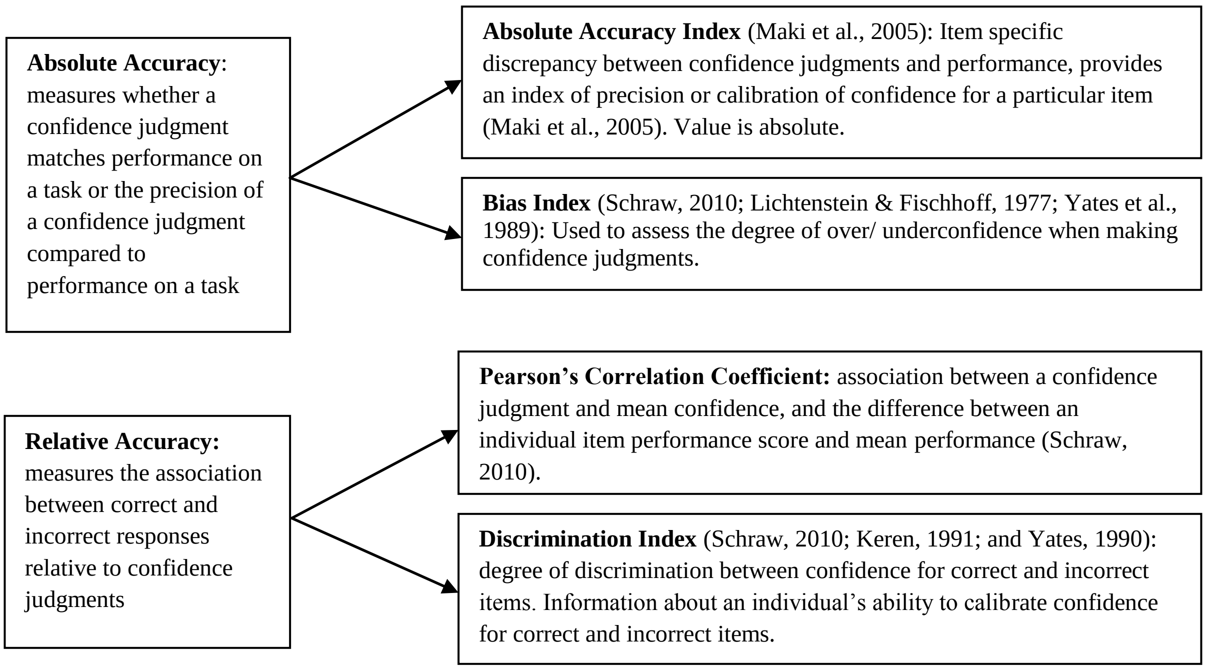
The two main ways to measure the goodness of fit between confidence judgments and performance is through examining absolute and relative accuracy (Schraw, 2010). Absolute accuracy measures whether a confidence judgment matches performance on a task or the precision of a confidence judgment compared to performance on a task. Relative accuracy measures the association between correct and incorrect responses relative to confidence judgments. Relative accuracy is typically represented by a correlational measure that reflects the association between confidence judgments and performance (i.e. an increase in performance should be matched with an increase in confidence to be well calibrated).

Absolute Accuracy. The Absolute Accuracy Index (Maki, Sheilds, Wheeler, & Zacchillo, 2005) and the Bias Index (Schraw, 2010; Lichtenstein & Fischhoff, 1977; Yates et al., 1989) were used to assess absolute accuracy. The Absolute Accuracy Index is a measure of the item specific discrepancy between confidence judgments and performance and thus provides an index of precision or calibration of confidence for a particular item (Maki et al., 2005). The second measure of absolute accuracy is the Bias Index (Lichtenstein & Fischhoff, 1977; Schraw, 2010) which is used to assess the degree of under/overconfidence when making confidence judgments.

Relative Accuracy. Relative accuracy was measured in the present study with Pearson's Correlation Coefficient and the Discrimination Index. Pearson's Correlation Coefficient represents the association between a confidence judgment and mean confidence, and the difference between an individual item performance score and mean performance (Schraw, 2010). The Discrimination Index (Schraw, 2010; Keren, 1991; and Yates, 1990) assessed the degree of discrimination between confidence for correct and incorrect items. This index provides information about an individual's ability to distinguish between confidence for correct and incorrect items. A summary of the performance calibration measures utilized in this study is presented in Figure 1.

Figure 1

Measures of Performance Calibration



Task Parameters that Influence Performance Calibration

Various factors can influence the accuracy of judgments about performance including question and answer format, question difficulty and delay.

Question and answer format.

There are aspects about the question-answer format that can influence how overconfident people appear (Klayman & Soll 1999). When confidence judgments are elicited after a question that requires a choice between two alternatives, there is only a slight bias toward overconfidence. Such questions resemble the format: What is the capital of Ontario? (a) Toronto (b) Ottawa. However, a question that asks people about a single estimate and then asks those people to rate their confidence around a margin of error surrounding that estimate, on a subjective scale, tends to elicit a large bias towards overconfidence. An example of this type of question is: How many people live in Canada? A correct answer would fall between 30 and 40 million, with the answer being approximately 35 million people. The current study elicited confidence judgments after questions that required both multiple choice answers and two alternative answers (i.e. Yes/No) in an attempt to limit a bias towards overconfidence.

Question difficulty.

Question difficulty can impact confidence ratings (Klayman & Soll 1999). When questions are easy people tend to be underconfident in their responses. When a set of questions are harder or harder questions are embedded within a set of questions, overconfidence becomes more pronounced. Gigerenzer and colleagues (1991) argued that overconfidence stems not from a pervasive cognitive bias but instead from the experimenters' tendency to choose "harder-than-normal questions". The phenomenon of easy items producing high accuracy and underconfidence, and difficult items producing low accuracy and overconfidence, has been termed the *hard-easy effect* (Juslin, Winman, & Olsson, 2000; Lichtenstein & Fischhoff, 1977). Kruger and Dunning (1999) provide evidence to suggest that people who are unskilled in particular domains suffer a dual burden: they make erroneous conclusions and their incompetence prevents them from having the metacognitive ability to realize their mistakes. Therefore, the harder people find questions, the more overconfident the ratings. The *hard-easy effect* has also been demonstrated with performance calibration in reading comprehension tasks (Lin, Zabrucky, & Moore, 2002). In the current study, items in the general knowledge task were categorized as hard and easy depending on the number of participants that answered the questions correctly in order to examine this *hard-easy* effect. Since children tend to be overconfident on difficult items, it is hypothesized that children in both groups will be better calibrated on easy items of the General Knowledge Task compared to hard items on the task, with children in the clinical ADHD group significantly more poorly calibrated on hard items.

Delay.

Delay and its effect on confidence judgments (i.e. concurrent or retrospective) has already been discussed above and the research indicates that performance calibration is typically more accurate following a delay (Kimball & Metcalfe, 2003; Nelson, Narens, & Dunlosky, 2004; Thiede, Anderson, & Therriauly, 2003; Thiede et al., 2005). In a child study, Schneider and colleagues (2000) found that JOL in children were more accurate after a two minute delay than immediately after each item. Briefly stated, *when* children are asked to make confidence judgments may play an important role in how confident they appear to be because delay may allow for the opportunity to engage in further metacognitive monitoring (Schraw, 2008). Therefore, the current study used both concurrent (confidence) and retrospective (effort and difficulty) judgments to gather more reliable measure of performance calibration. Finally, methodological considerations for performance calibration measures were addressed and measures were assessed for their experimental and clinical utility.

Summary and Hypotheses

Performance calibration has been shown to be impaired in ADHD in the following ways: Positively biased or overconfident self-perceptions (Evangelista et al., 2008; Fabiano et al., 2015; Fefer et al., 2015; Gerdes et al., 2003; Hoza et al., 2000, 2001, 2004; Jia et al., 2015; Scholtens et al., 2012; Owens & Hoza, 2003), lack of post-error slowing and problems with error detection (Balogh & Czobor, 2014; O'Connell et al., 2009; Schachar et al., 2004) and lack of reliance on self-report of symptoms in ADHD (CADDRA, 2011; Pliszka, 2007; NHMRC, 2009; Taylor et al., 2004). Performance calibration or the accuracy between actual performance and an individual's judgment of his or her performance provides an alternative method from the PIB literature for operationalizing self-monitoring of performance in children (Desoete & Roeyers, 2006; Newman, 1984; Weller et al., 2012). No studies have compared performance calibration and under/overconfidence across domains in children with and without ADHD. The general finding in performance calibration studies is that confidence exceeds accuracy of judgments (Klayman & Soll, 1999). The purpose of the current study was to compare performance calibration paradigms from the cognitive science literature between a community sample of TD children and a clinical sample of children with ADHD aged 8-12 years-old. Based on the emerging literature which suggests poor performance calibration in children with ADHD, we expected to obtain the following findings in the novel measures utilized in this study:

Hypothesis 1: Children with ADHD were predicted to be less accurate and more confident than the TD group on absolute accuracy measures. This difference was expected on the three performance calibration tasks used in this study: the General Knowledge Task, Emotion Recognition Task, and the Child Social Situation Questionnaire.

Hypothesis 2. Children with ADHD were predicted to have significantly poorer performance calibration on relative indices compared to the TD sample. This was expected on the three performance calibration tasks used in this study: the General Knowledge Task, Emotion Recognition Task, and the Child Social Situation task.

Hypothesis 3. Based on the literature finding the *hard-easy effect* (Juslin et al., 2000; Lichtenstein & Fischhoff, 1977), it was expected that all children would be overconfident on difficult items but underconfident or better calibrated on easy items of the General Knowledge Task. Given that children in the clinical ADHD group were predicted to be less well calibrated overall (hypotheses 1 and 2), it was expected that while both groups would be overconfident on difficult items, the clinical group may demonstrate more overconfidence on difficult items and underconfidence on easy items relative to the control group. The General Knowledge Task was chosen to evaluate the *hard-easy effect* because questions on this task were intended to range in difficulty.

Hypothesis 4. Based on the decision-making literature (Stanovich & West, 2000; Wolfe & Grosch, 1990), it was expected that the Absolute Accuracy Index would be significantly negatively associated with estimated intelligence and EFs. That is, low absolute accuracy (better calibration) would be associated with better cognitive abilities. The Absolute Accuracy Index was chosen to represent performance calibration here since absolute miscalibration, and not directionality of miscalibration (i.e. Bias Index scores) has been implicated in the decision-making literature.

Hypothesis 5: Since the diagnostic criteria for ADHD includes avoidance of effortful tasks in its symptom criteria for inattention, it was predicted that participants in the clinical ADHD group would rate the performance calibration tasks as more difficult and effortful than the control group and that parents of children in the ADHD group would rate their child's performance calibration lower than the TD group. In order to determine whether confidence was calibrated with perceived difficulty, a method similar to the discrimination indices for performance calibration was used. In other words, if a task is perceived to be difficult, then this should be reflected with a lower mean confidence rating of that task.

Hypothesis 6. Since confidence for accuracy of emotion recognition is a new area of research, performance calibration across the five emotions presented in the ERT were explored. It was hypothesized that children with ADHD may have worse performance calibration than the TD group overall. The Bias Index was chosen to represent performance calibration here because the directionality of miscalibration with regard emotions was of interest for examination (i.e. over/underconfident across groups and emotions).

Hypothesis 7. Confidence ratings for accuracy in identifying ambiguously positive and negative social situations in the Child Social Situations Questionnaire were explored. Since children with ADHD have been found to be impaired in the detection and encoding of external cues (Andrade et al., 2012) and may respond more negatively than do typically developing children (Mattys et al., 1999), it was expected that children in the clinical ADHD group would be better calibrated on the ambiguous negative social situations and less well calibrated on the ambiguous positive social situations compared to the TD group. The Absolute Accuracy and Bias Indices were chosen to represent performance calibration here because the directionality of miscalibration and absolute miscalibration with regard each social situation was of interest for examination.

Method

Participants

A sample of 81 8 to 12-year-old children [*M* = 119.89 months (or 9 years, 11months), *SD* = 14.92 months, 26 females] from the Greater Toronto Area participated in this study. Based on effect size calculations from previous research measuring performance calibration, in order to have an anticipated effect size of Cohen's d = .6, approximately 80 participants (40 per group) were necessary for 80% power to detect significant group differences in confidence ratings about performance (alpha=.05). There were two groups of children in the sample; a clinical ADHD sample consisting of children with an ADHD diagnosis and a typically developing (TD) control sample of children without ADHD. Children were included in the study if they met the following inclusion criteria: 1) Between the ages of 8-12 years-old, 2) Have a prior diagnosis of ADHD and meet diagnostic criteria for ADHD based on the Computerized Diagnostic Interview Schedule for Children – Parent Version (C-DISC; Shaffer et al., 2000; Fisher, Lucas, Lucas, Sarsfield, & Shaffer, 2006) to be in the clinical ADHD group, 3) Must not have a prior diagnosis of ADHD and must not meet diagnostic criteria for ADHD on the C-DISC nor the CBCL to be in the TD group, and, 4) Must have Borderline or above intellectual functioning (IQ > 70) and not have an Autism Spectrum Disorder (ASD). The diagnostic algorithm for ADHD on the CDISC comprises symptom counts, pervasiveness across setting and level of impairment required for the DSM-IV diagnosis as described in the introduction while the CBCL provides a T-score for the ADHD scale with a Tt-score > 70 representing the clinically significant range for the DSM-V diagnostic criteria. Children meeting these criteria were invited to participate in the study. Since the age distributions in each group were not normal, an Independent Samples Mann-Whitney U Test was executed to determine if there was a significant difference in the distribution of age (in months) across the TD and clinical ADHD groups. There was no significant difference in age distribution across the two groups (p = .606).

Participants in the TD sample were recruited by advertising in a local community (arenas and public library) who indicated interest in being contacted to participate in research. The TD group consisted of 15 females and 27 males aged 8-12 years-old [*M* = 120.43 (10 years, 0 months), *SD* = 14.73)]. Parents identified 78.6% children as Caucasian, 16.7% as having a mixed background, 2.4% as Latin American, and 2.4% as other. Ethnicities represented in children identified as mixed background included Caribbean, African, European, Arab, and East Asian. No parents reported any of the TD children to be taking any psychotropic medication. One child (2.4%) was identified as having a Learning Disorder, with impairments in reading and spelling. One child (2.4%) was indicated as having some articulation difficulties.

The clinical ADHD sample was recruited from an outpatient mental health service, private psychological practices, and online through the Canadian ADHD Resource Alliance (CADDRA) and the Centre for ADHD Awareness, Canada (CADDAC). The clinical ADHD group consisted of 11 females and 28 males aged 8-12 years-old [M = 119.31 months (9 years, 11 months), SD = 15.29 months]. Parents identified approximately half of the children as Caucasian (48.7%), 30.8% as mixed background, 17.9% as other, and 2.6% as Latin American. Ethnicities represented among children identified as mixed background included Caribbean, Indochinese, Czech and Slovak, and Aboriginal.

Parents in the clinical ADHD group reported that approximately half of the children were taking psychotropic medication (51.3%), and the remaining children were not taking any psychotropic medication (48.7%). Of the children taking psychotropic medication 19 were taking psycho-stimulants and one was taking a non-psychostimulant: This child was taking a selectiveserotonin reuptake inhibitor to treat anxiety in addition to a psycho-stimulant for attentional difficulties. In addition one child was taking homeopathic remedies for concentration. An Independent Samples Mann-Whitney U Test was conducted to determine if there were significant differences between IQ, set-shifting and inhibition between children in the clinical ADHD group based on medication status. There were no significant differences in median scores between the medicated an unmedicated participants in these group on these measures ($Z = -.20$, $p = .84$; $Z = -1.04$, $p = .30$; and $Z = -1.15$, $p = .25$, respectively).

Ten caregivers from the clinical ADHD group indicated that their child had a Learning Disorder (25.6% of the group). Two children were indicated as having a Language Impairment (5.1%). Eighty-seven participants were tested in total. Five participants were excluded from the clinical ADHD group. Four participants were excluded because they did not meet ADHD diagnostic criteria on the CDISC or the CBCL and one participant was excluded because his IQ score was below the inclusion criterion of 70. One participant was excluded from the TD group because he met criteria for ADHD on the CDISC. The total sample size was 81.

Parent demographics.

In the TD group, 39 mothers and three fathers completed the study with their child. The marital status of parents in this group included 38 married parents, three common-law marriages, and one single parent. Level of mother’s education in this group included one mother who completed high school, three mothers who completed some post-secondary education, and 37 mothers who completed post-secondary education.

In the clinical ADHD group, 33 mothers, three fathers, one grandmother, one adoptive father, and one adoptive mother completed the study with their child. The marital status of parents in this group included 19 married parents, two common-law marriages, 10 single parents, 4 divorced parents, one engaged parent, two separated parents, and one parent who did not indicate their marital status. Level of mother's education in this group included one mother who completed up to Grade 8, four mothers who completed high school, 10 mothers who completed some post-secondary education, 23 mothers who completed post-secondary education, and one mother who did not disclose her highest level of education.

Measures

Child measures.

Intellectual and EF abilities.

Kaufman Brief Intelligence Test, Second Edition (KBIT-2; Kaufman & Kaufman, 2004). The KBIT-2 is a brief individualized test that consists of three subtests that measure crystallized and fluid intelligence. Crystallized intelligence was assessed verbally with a receptive and an expressive vocabulary tasks that did not involve reading or spelling demands. The first verbal subtest, Verbal Knowledge, is composed of 60 items that require the participant to choose which of five-to-six pictures depict a word spoken by the examiner. The second verbal subtest, Riddles Subtest, has 48 items and requires the participant to solve a verbal riddle with a one-word answer (or identification of a picture for younger ages). Fluid intelligence was assessed nonverbally with a matrix reasoning task. The Matrices Subtest has 46 multiple choice items that require participants to identify which picture best completes a matrix.

Split-half reliability for the Verbal and Composite IQ scores are high in the mid .80s to mid .90s and the mean reliability for the Verbal Scale score was .91 and .93 for the Composite IQ score (Kaufman & Kaufman, 2004). The split-half reliability for the Nonverbal Scale fell in the .80s and .90s for most age groups. Test-retest reliability ranged from .88 to .93 on the Verbal score, from .76 to .89 for the Nonverbal scale, and from .88 to .92 for the IQ composite. Validity was also demonstrated with multiple intelligence tests, such as the WASI verbal components with a correlations ranging from .80 to .86 and correlations between .62 and .80 across nonverbal components, and from .81 to .90 for IQ composite/Full Scale IQ scores. A composite, non-age corrected, score of intellectual abilities was obtained by summing the standardized z-scores of each raw score (verbal and nonverbal tests). Higher scores are indicative of higher intellectual abilities.

Executive functioning. The Trailmaking Test (Reitan, 1958) was used to measure set-shifting or mental flexibility. Set-shifting is a cognitive task that requires one to display flexibility when there are changing rules/schedules of reinforcement in the environment (Strauss et al., 2006). The Trailmaking Test consisted of two parts with practice items for both parts. Part A required participants to connect 25 numbers circled in numeric order with a pencil. Part B consisted of 13 circled numbers and 12 circled letters, and the participant was instructed to alternate between letters and numbers 1 to A, A to 2, 2 to B and so on), until all of the circled numbers and letters were exhausted. Two participants had extreme outlying scores on Trailmaking that were imputed with the next most extreme value within their group. The dependent measure on this task was the completion time on Part B minus the completion time on Part A. By subtracting Part A from Part B, processing speed is controlled for, and the time remaining is a measure of set-shifting alone. Higher scores indicated worse set-shifting.

The Stroop Task (Stroop, 1932) was used to measure interference control, a type of inhibition (Friedman & Miyake, 2004). Interference control refers to the ability to filter out irrelevant information and to select relevant information. There were three different conditions, each with 24 items arranged in a 4 x 6 matrix: a word reading condition, a colour naming condition, and an interference condition. In the word reading condition, participants were presented with a chart of 48 words that named four colours (red, green, blue, yellow) displayed in a matrix of six columns and eight rows. Participants were asked to read the words as quickly as possible without making any errors. In the colour naming condition, participants were presented with a chart of 48 patches of red, green, blue, and yellow colours displayed in a matrix of six columns and eight rows. Participants were asked to name the colours as quickly as possible without making any errors. In the interference condition, participants were presented with a chart of 48 words displayed in a matrix of six columns and eight rows. In this condition, the colour naming words (red, green, blue, yellow) appeared in a different colour ink (red, green, blue, yellow) than the colour the word named. For example, the word "red" appeared in the colour yellow. Participants were asked to name the colour as quickly as possible without making any errors. The interference condition was the most difficult of the three conditions because participants must inhibit the competing modality of naming the words. The dependent variable was the total naming time for the interference condition minus the total naming time for the colour naming condition (Strauss et al., 2006). Higher scores indicated lower inhibition.

Measures assessing performance calibration.

1) General Knowledge Task (GKT). This task was developed for the current study and included 36 questions about animals, drinks, planets, and other general topics. The questions on this task were modified from a free online question bank for children of various ages (www.free-for-kids.com). For each question, there were four options that the participant chose from, followed by a confidence rating for each item. See Appendix B for task instructions, items, and performance on this task. Prior to the current study, this task was piloted with 49 children aged eight to 11 to ensure the age-appropriateness of test items. Higher scores indicated better performance on this task.

Hard vs. easy items on the GKT. In order to distinguish between hard and easy items on the GKT, items for which at least 75% of respondents answered correctly were denoted “easy” and items for which less than 75% answered correctly were denoted “hard.” Based on these cut-offs for the current sample, there were 12 hard items and 24 easy items.

2) Emotion Recognition Task (ERT). This task was developed for the current study using the NIMH Child Emotional Faces Picture Set (Egger et al., 2011). The NIMH Child Emotional Faces Picture Set is a large, publically available, stimulus set with validated high quality colour images of children’s faces depicting fearful, angry, happy, sad, and neutral emotions. This task consisted of 40 items, with each emotion represented eight times, and of these eight representations, four were female and 4 were male faces. For each item, participants were shown an image of a child’s face at the top of the page and the five emotions were listed at the bottom of the page. Participants first selected which emotion matched the image and then rated his/her confidence for each item. See Appendix C for task instructions, items, and performance on this task. Higher scores indicated better performance on this task.

3) Modified Child Social Situation Questionnaire (CSSQ; Andrade et al., 2012). This measure was adapted to examine children’s performance calibration when interpreting social situations. One practice and eight social information processing vignettes were presented to each participant. On the original CSSQ, children were instructed to pretend that they are the protagonist in each story, were read each story aloud, and then orally answered three questions designed to evaluate different aspects of each child’s social information processing: 1.What happened in the story? 2. How could you tell whether this was a nice way to act or a mean way to act? 3. What could you say or do if this happened to you? Tell me as many ways as you can.

There were four vignettes that had an ambiguous-positive outcome and four vignettes that had an ambiguous-negative outcome.

For the current study, the first question was adapted so that the child's accuracy in interpreting the social situation could be evaluated. Thus, after explaining what happened in the story, participants were asked 1. Do you think that something good/positive happened in the story? (Y/N), 2. How confident/sure are you? (on the confidence scale below), 3. Do you think that something bad/negative happened in the story? (Y/N), and 4. How confident/sure are you? The CSSQ was audio recorded and participant answers were transcribed. See Appendix D for the modified CSSQ situations, instructions, scoring, and performance on this task. Participants were allotted a point if they correctly identified whether the situation was positive or negative based on the vignette descriptors. Higher scores indicated better performance on this task.

Indicators of performance calibration across all three tasks.

While participants completed the General Knowledge, Emotion Recognition, and Child Social Situation tasks, they also made concurrent judgments about their confidence in the accuracy of their answers. The judgments were made with a confidence rating scale that asked each participant to judge how confident they were in the accuracy of their answers, on a Likert scale of 0 – Do not know, just guessing; to 10 – I'm definitely right. The confidence rating scale was based on previous rating scales such as those used in the Feeling of Rightness measure (Thompson, Prowse Turner, & Pennycook, 2011), measures of confidence in the correctness of answers (Koriat & Ackerman, 2010), and judgments of confidence measures (Schraw, 2009; Weller et al., 2012). See Appendix E for the confidence rating scale. Multiple outcome measures of performance calibration were employed in the current study.

The Absolute Accuracy Index measures item specific discrepancy and is represented by the following equation:

$$1/n \, \Sigma \, (c_i - p_i)^2$$

where c_i is the confidence rating for each item out of 10 (i.e. a confidence of 5 = 5/10 = .5), p_i is the performance score for the same item (i.e. correct is 1 and incorrect is 0), and n is equal to the number of confidence judgments (Maki et al., 2005). Each deviation score between the participant's confidence score and their performance score is squared so that absolute values are obtained and then divided by the number of items. Scores that are close to zero correspond to good performance calibration, and large positive scores toward the upper end of the scale reflect poor performance calibration.

The Bias Index (Schraw, 2010) was used to assess the degree of over/underconfidence when making confidence judgments. Lichtenstein and Fischhoff (1977) and Yates and colleagues (1989) also apply this measure to the calculation of over/underconfidence in probability judgments by subtracting the mean percentage correct from the mean percentage confidence judgment. Lichtenstein and Fischhoff (1977) and Yates and colleagues' (1989) measure produces the same difference score with positive values corresponding to overconfidence and negative values corresponding to underconfidence identical to the Bias Index.

The Bias Index (Schraw, 2010) is represented by the following equation:

$$1/n \, \Sigma \, (c_i - p_i)$$

where c_i is the confidence rating out of 10 for each item, p_i is the performance score for the same item (i.e. correct is 1 and incorrect is 0), and n is equal to the number of confidence judgments.

The Bias Index differs from the Absolute Accuracy Index in that it does not square the deviation between the confidence judgment and performance so that the value of the index can be positive or negative. When the judgment of confidence is high but the performance score is low (i.e. incorrect), this demonstrates overconfidence and the index is positive and above zero. When the judgment of confidence is low but the performance score is high (i.e. correct), this demonstrates underconfidence and the index is negative and below zero. When confidence is matched with performance (such as low confidence and low performance), the individual is well calibrated and the Bias Index score is close to or at zero. Thus, while the Absolute Accuracy Index ranges from well-calibrated (zero) to miscalibrated (above zero, with higher scores representing worse performance calibration), the Bias Index ranges from underconfident (miscalibrated and a negative value below zero) to well-calibrated (zero) to overconfident (miscalibrated and a positive value above zero).

Pearson's Correlation Coefficient measured the association between a confidence judgment and mean confidence, and the difference between an individual item performance score and mean performance (Schraw, 2010). The formula for *r* using standardized variables is:

$$\Sigma Z_x Z_y/(n - 1),$$

$$\text{where } Z_x = \Sigma (X_1 - \overline{X})/s$$

A positive correlation indicates that deviation scores are in the same direction (as confidence scores go up, so does performance) while a negative relationship indicates that deviation scores are in the opposite direction (as confidence scores go up, performance goes down or vice versa). A correlation of zero indicates that there was no association between accuracy and confidence judgments (i.e. that there was no slope when confidence judgments were plotted against performance outcomes).

The Discrimination Index (Schraw, 2010; Keren, 1991; and Yates, 1990) assessed the degree of discrimination between confidence for correct and incorrect items with the formula:

$$1/n_c\ \Sigma\ (c_{i,\ correct})\ -\ 1/n_i\ \ \Sigma\ (c_{i,\ incorrect})$$

where $c_{i,\ correct}$ is the confidence for each correct item, $c_{i,\ incorrect}$ is the confidence for each incorrect item, n_c is the total number of items answered correctly, and n_i is the number of items answered incorrectly. This measure differs from absolute accuracy and bias measures in that it monitors one's confidence across items rather than performance on specific items since it is computed based on aggregate data. When the score on this index is positive, the individual is more confident about correct versus incorrect items. When the score on discrimination is negative, the individual is more confident about incorrect items. Higher values represent higher confidence ratings for correct items and lower confidence ratings for incorrect values. A score of zero means that individuals did not discriminate with their confidence judgments across correct and incorrect items.

Task difficulty and cognitive effort. Task difficulty and cognitive effort were measured after each of the child tasks in the current study as a retrospective indicator of performance calibration. See Appendix F for the items and the pictorial Likert scale. For *task difficulty*, participants were asked to indicate how hard they found the task with a pictorial Likert scale ranging from not hard to very hard (scale from Schäfer, 2011). A stick figure was depicted carrying a back-pack that got larger and larger toward the end of a 5 point scale. Participants marked the picture that represented how difficult they found the preceding task. For *cognitive effort*, participants were asked to indicate how hard they tried on each task with a Likert scale ranging from 1 – I did not try at all to 5 – I tried very hard.

Reliability and validity was assessed for the clinical ADHD and TD groups separately and jointly for all of the performance calibration measures. Reliability of the confidence ratings was assessed using Cronbach's Alpha for the General Knowledge Task, Emotion Recognition

Task, and the Child Social Situation Task. (See Table 1). Both the GKT and ERT had exceptional reliability, while the CSSQ was lower yet acceptable.

Table 1

Reliability of Confidence Ratings for the GKT, ERT, and CSSQ Task

Measure	Cronbach's Alpha - whole sample	Cronbach's Alpha - TD group	Cronbach's Alpha - Clinical sample
General Knowledge Task - Confidence Ratings Items 1 – 36	.90	.87	.91
Emotion Recognition Task - Confidence Ratings Items 1 – 40	.96	.93	.98
CSSQ Task - Confidence Ratings Items 1 – 8	.68	.70	.47

Note: CSSQ = Child Social Situation Questionnaire

Parent measures.

Demographics Questionnaire. Parents/guardians completed a demographics questionnaire about themselves and their child. Questions included information about the child such as age, country of birth, medications, and how well the parent perceives his/her child to monitor task performance with the question "This study is about metacognition, or how well your son/daughter can monitor his/her performance on tasks. How would you rate your child's metacognition on tasks?" On this question, parents complete a Likert scale ranging from 1 – poor tracking of accuracy of work to 7 – excellent tracking of accuracy of work. Parents/guardians provided information about themselves such as ethnic belonging, marital status, and socioeconomic indicators such as highest level of education and occupation. See Appendix G for demographics questionnaire.

Computerized Diagnostic Interview Schedule for Children – Parent Version (C-DISC; Shaffer et al., 2000; Fisher et al., 2006). The C-DISC is a structured interview designed to assess DSM-IV psychiatric disorders, symptoms, and level of impairment in children and adolescents aged six to 17 years. For the purposes of the current study, the change in DSM-IV criteria to the DSM-V criteria is minimal and involves the age of onset which is not relevant here. Parents answered questions about whether their child has experienced a specific symptom over the past year and then answered follow-up questions when items were endorsed. The Attention/Deficit-Hyperactivity Disorder subscale, the Oppositional Defiant Disorder (ODD) subscale, and the Conduct Disorder (CD) subscale were administered to parents by trained graduate clinical psychology students. This measure was used to confirm and rule-out ADHD diagnoses for the current study. The dependent variables used from this task for inclusion/exclusion in the study and sample characteristics were the number of ADHD symptoms and impairment, ODD number of symptoms in past year and past year impairment and CD number of symptoms in last year and past year impairment.

Child Behavior Checklist (CBCL, Achenbach & Rescorla, 2001; Achenbach, 2013). This checklist was completed by parents and is commonly used to detect different emotional and behavioural problems in children and adolescents. It was designed to be used with children aged six to 18 years-old. Questions were scored on a three-point Likert scale ranging from 0 = Absent, 1 = Occurs Sometimes, and 2 = Occurs Often. In order for an item to be endorsed in must have been present in the past six months. The inter-interviewer and test-retest reliability of item scores were high with intraclass correlation coefficients (ICC) ranging from .93 to 1.00. The CBCL is made up of eight syndrome scales that group into the two higher factors of internalizing and externalizing; anxious/depressed, depressed, somatic complaints, social problems, thought problems, attention problems, rule-breaking behaviour, and aggressive behaviour. There are also six DSM-5-oriented scales consistent with the DSM diagnostic categories of: affective problems, anxiety problems, somatic problems, ADHD, oppositional defiant problems (ODP), and conduct problems (CP). The CBCL was administered in order to characterize the sample and to support inclusion/exclusion into the clinical and TD groups through the following dependent variables: the ADHD T-Score, ODP T-Score, and CP T-Score.

Impairment Rating Scale (IRS; Fabiano et al., 2006). The IRS is a seven-item measure of child functioning and need for treatment in several developmentally important areas. The domains assessed include impairment in peer relationships, sibling relationships, parental relationships, academic progress, self-esteem, general family functioning, and overall severity of problem behaviour. Parents were presented with a visual-analogue scale and were asked to indicate the level of problems in each area on a 0 - no problems/no need for treatment to 6 - severe problems/definitely needs treatment metric. The IRS's reliability and validity were assess by Fabiano and colleagues (2006). The IRS has a cross-informant reliability of .78 and one-year temporal stability correlations ranging from .54 to .76. Convergent validity has been shown with ADHD symptoms on the DISC and IRS with correlations ranging from .58 to .79. The mean impairment rating across each item was used in the present study to characterize the level of impairment experienced by each participant. A higher score indicated more impairment.

Barkley Deficits in Executive Functioning Scale – Children and Adolescents (BDEFS-CA Short Form; Barkley, 2012). The BDEFS- CA short form is a 20-item empirically based tool that was used to evaluate child and adolescent executive functioning. The BDEFS-CA measured some of the capacities involved in time management, organization and problem solving, self-restraint, self-motivation, and self-regulation of emotions. This measure has been found to be both reliable and valid (Barkley, 2012), with internal consistency of .95 for the 20-item form and convergent validity with the BRIEF summary scores ranging from .83 to .92. The short parent-report form was administered to parents in order to characterize each participant's executive functioning. An overall score was derived and higher scores indicated more deficits in executive functioning.

The Inventory of Callous-Unemotional Traits – Parent Report (ICU; Essau, Sasagawa, & Frick, 2006). The ICU is a 24-item questionnaire designed to assess callous and unemotional and was included in the current study since oppositional defiant bahaviours and ADHD often co-occur (Tannock, 1998). The ICU has three subscales: Callousness, Uncaring, and Unemotional. Parents completed this measure in order to characterize psychopathology of the children in this sample. An overall score was derived and higher scores indicated more callous-unemotional traits.

Procedure

For each testing session, two research assistants met with each child and his/her parent. Informed consent and assent were first obtained to participate in the study. One examiner administered the measures to the child, and the other examiner administered the parent measures and completed the diagnostic computerized interviewed with the parent. Examiners were sensitive to reading and language difficulties, and all materials were read to participants. Study completion time ranged from 90 to 120 minutes and each participant received a small honorarium of $20. A smaller project was also part of this project and included one additional measure not included here.

The child measures were counterbalanced such that half of the battery was presented first or last dependent on whether the participant had an odd or even participant identification number. Participants with odd numbers received Version A of the battery and participants with even numbers received Version B of the battery (See Appendix A for checklist version and ordering). The medians on the metacognition tasks were compared using the Independent Sample Mann-Whitney U Test to determine whether there were any differences in accuracy and confidence for those who completed the metacognition tasks at the beginning or the end of the battery (i.e. Version A or Version B). The median General Knowledge Task accuracy scores did not differ across the different versions ($Z = -.31$, $p = .76$), but the Emotion Recognition Task and the Child Social Situation accuracy scores did differ ($Z = -2.45$, $p = .01$ and $Z = -2.57$, $p = .01$, respectively). The median confidence ratings based on version did not differ across these three tasks ($Z = -.07$, $p = .95$, $Z = -1.16$, $p = .25$, and $Z = -.68$, $p = .49$, respectively). These findings suggest that those who completed these tasks later in the battery had lower accuracy scores (on the Emotion Recognition and Child Social Situation tasks), suggesting some fatigue effects for accuracy but not for confidence ratings.

Results

Statistical Analyses

All statistical analyses were conducted using SPSS version 23.0. The significance level for research questions was set at the standard $p < .05$. The normality of each of the variables measured was tested using the Shapiro-Wilk test of normality as well as visual inspection of histogram and Q-Q plots. Some of the distributions were not normal and therefore nonparametric tests such as Spearman Rank Order Correlations and Independent Samples Mann-Whitney U Tests were employed throughout. When ANOVAs were used, assumptions were checked and met.

Missing data. One parent did not complete the CDISC, and the CBCL was used to determine inclusion in the TD group. Three data points were missing in the TD group. Missing data were imputed based on group means, and this was done for items on one CDISC, one CBCL, and one BDEFS.

Descriptive Statistics by Group

CDISC diagnosis frequencies.

Based on symptom level and impairment, the CDISC generates a diagnostic designation for each participant for ADHD, ODD, and CD based on the DSM-IV criteria. The diagnostic designations for each group are presented in Table 2. All participants in the clinical ADHD group met diagnostic criteria for ADHD with 56.4% also meeting criteria for ODD, and 12.8% meeting criteria for CD. None of the participants in TD met criteria for ADHD as part of the inclusion criteria for the group, however 9.5% of this group had an intermediate ADHD classification (no impairment), 7.1% had an intermediate ODD classification (no impairment), and none had any CD classifications.

Table 2

CDISC Diagnoses by Group

	TD Group			Clinical ADHD Group		
Measure	*Positive*	*Intermediate*	*Negative*	*Positive*	*Intermediate*	*Negative*
ADHD Diagnosis	0	4	38	39	0	0
ODD Past Year Diagnosis	0	3	39	22	12	5
CD Past Year Diagnosis	0	0	42	5	3	31

Note: ODD = Oppositional Defiant Disorder, CD = Conduct Disorder

CDISC, BDEFS, ICU, CBCL, and IRS group differences.

Table 3 provides the medians and standard deviations for each group on the dependent measures for the CDISC, BDEFS, ICU, CBCL, and IRS. Group medians were statistically compared and the results indicated that the TD group was significantly different from the clinical group on all measures except for the ICU indicating that the level of unemotional callous traits did not differ between groups. The clinical ADHD group had significantly higher symptom counts and impairment on the CDISC, CBCL, and IRS for ADHD, ODD and CD.

Table 3

Clinical Demographics by Group

Measure	TD Group		Clinical ADHD Group		Independent Samples Mann-Whitney U Test	
	Mdn	*SD*	*Mdn*	*SD*	*Statistic*	Z-score
CDISC – ADHD Number of Symptoms /23	1	1.76	15.00	2.92	.000	-7.81**
CDISC – ADHD Impairment	.00	1.74	12.00	3.02	7.5	-8.07**
CDISC – ODD Past Year Number of Symptoms /12	3.00	3.41	9.00	2.56	211.50	-5.77**
CDISC – ODD Past Year Impairment	.00	1.53	10.00	5.02	128.00	-7.09**
CDISC – CD Past Year Number of Symptoms /26	.00	.41	3.00	3.21	192.00	-6.29**
CDISC – CD Past Year Impairment	.00	.00	.00	6.86	609.00	-3.47**
BDEFS Summary Score	28.82	6.65	54.00	11.05	58.00	-7.20**
ICU Total Score	26.50	3.75	26.00	4.51	757.50	-.58
CBCL – ADHD T Score	50	2.69	66	7.18	55.00	-7.36**
CBCL – Oppositional Defiant Problems T Score	51	3.25	63.68	8.40	158.00	-6.32**
CBCL – Conduct Problems T Score	50	4.32	60.42	8.36	296.50	-5.06**
IRS	0	.67	3.43	1.43	48.50	-7.40**

*$p < .05$, **$p < .01$.
Note: CDISC = Computerized Diagnostic Interview Schedule for Children, BDEFS = Barkley Deficits in Executive Functioning Scale, ICU = Inventory of Callous and Unemotional Traits, CBCL = Child Behavior Checklist

Cognitive abilities.

Group medians on cognitive abilities are presented and compared in Table 4 using Independent Samples Mann-Whitney U Tests. There were significant median differences between the groups with respect to the fluid/non-verbal intelligence and Trailmaking Part B minus Part A. Therefore, the clinical ADHD group had lower levels of non-verbal intelligence and took longer to shift sets as compared to the TD group.

Table 4

Group Differences in Cognitive Abilities

	TD Group		Clinical ADHD Group		Independent Samples Mann-Whitney U Test	
Measure	*Mdn*	*SD*	*Mdn*	*SD*	*Statistic*	Z-score
KBIT –Estimated IQ Composite (age-corrected)	109.00	8.42	103	15.23	660.00	-1.50
KBIT – Verbal Raw Score	62.00	9.85	61.50	12.53	775.50	-.41
KBIT – Nonverbal Raw Score	32.00	3.82	30.50	6.18	604.50	-2.03*
KBIT – Composite z-score	.08	1.11	-.30	1.94	676.00	-1.35
Trailmaking Part B minus Part A Time	74.50	43.69	103.5	71.90	570.00	-2.35*
Stroop Interference Score	37.50	15.60	49.00	23.98	630.00	-1.62

*$p < .05$

Note: KBIT = Kaufman Brief Intelligence Test

Performance calibration.

Overall Accuracy and Confidence. Group medians and standard deviations for the total score on each of the performance calibration tasks are presented in Table 5 along with the mean confidence ratings. Independent Samples Mann-Whitney U tests did not indicate median differences between the two groups' GKT score, ERT score, or the CSSQ Task score. There were no median group differences in terms of the GKT mean confidence rating but median group differences were found on the ERT mean confidence rating and the CSSQ Task mean confidence rating indicating that the clinical ADHD group had significantly higher confidence ratings than the TD group on these tasks.

Table 5

Group Differences in Performance and Confidence on Metacognition Tasks

	TD Group		Clinical ADHD Group		Independent Samples Mann-Whitney U Test	
Measure	*Mdn*	*SD*	*Mdn*	*SD*	*Statistic*	*Z-value*
General Knowledge Task Score /36	30.00	3.01	30.00	3.81	779.50	-.38
Emotion Recognition Task Score /40	34.00	2.42	34.00	2.58	788.00	-.30
CSSQ Task Score /8	8.00	1.27	8	1.04	800.50	-.20
General Knowledge Task Mean Confidence Rating	8.50	.77	8.56	1.22	746.50	-.69
Emotion Recognition Task Mean Confidence Rating	9.04	.80	9.70	1.05	416.00	-3.82**
CSSQ Task Mean Confidence Rating	8.75	.92	9.63	.65	429.50	-3.70**

**p* < .05, ***p* < .01
Note: CSSQ = Child Social Situation Questionnaire

Analyses by Hypothesis

Hypothesis 1. It was hypothesized that children with ADHD would be less accurate absolutely and more overconfident than the TD group. This difference was expected on the three performance calibration tasks used in this study: the General Knowledge Task, Emotion Recognition Task, and the Child Social Situation task. Group median differences on the absolute accuracy performance calibration indices (Absolute Accuracy and Bias) were compared across groups for each of the GKT, ERT, and CSSQ Task with the Independent Samples Mann-Whitney U Test in Table 6. Significant median differences were found for the ERT Bias and the CSSQ Task Bias Index. The TD group was significantly less overconfident than the clinical ADHD group on the ERT but was significantly underconfident on the CSSQ with the clinical ADHD group being well calibrated.

Table 6

Group Differences on the Absolute Accuracy Performance Calibration Indices

	TD Group		Clinical ADHD Group		Independent Samples Mann-Whitney U Test	
Measure	*Mdn*	*SD*	*Mdn*	*SD*	Statistic	Z-value
GKT Absolute Accuracy Index	.11	.06	.13	.76	650.50	-1.59
GKT Bias Index	.01	.10	.03	.12	755.50	-.60
ERT Absolute Accuracy Index	.13	.05	.13	.07	707.00	-1.06
ERT Bias Index	.04	.10	.09	.12	565.50	-2.40*
CSSQ Task Absolute Accuracy Index	.06	.09	.03	.13	689.50	-1.23
CSSQ Task Bias Index	-.05	.15	.00	.14	519.50	-2.84**

*$p < .05$, **$p < .01$

Note: GKT = General Knowledge Task, ERT = Emotion Recognition Task, CSSQ = Child Social Situation Questionnaire

Hypothesis 2. Children with ADHD were predicted to have significantly poorer performance calibration on relative accuracy indices compared to the TD sample on the General Knowledge Task, Emotion Recognition Task, and the Child Social Situation task. The clinical ADHD group was thought to have a lower association between confidence judgments and performance than the TD group (i.e. an increase in performance should be matched with an increase in confidence to be well calibrated, but less so for the clinical ADHD group). Group median differences on the relative accuracy performance calibration indices (Relative Accuracy, and Discrimination) were compared across groups for each of the GKT, ERT, and CSSQ Task with the Independent Samples Mann-Whitney U Test in Table 7. Significant median differences were found for the ERT Relative Accuracy and Discrimination Indices. On the ERT, the TD group showed a significantly higher association in confidence judgments relative to the increase in performance outcomes (i.e. confidence aligned with performance outcomes) and was more discriminate in confidence ratings than the clinical ADHD group such that confidence ratings were higher for correct versus incorrect items.

Table 7

Group Differences on the Relative Accuracy Performance Calibration Indices

	TD Group		Clinical ADHD Group		Independent Samples Mann-Whitney U Test	
Measure	*Mdn*	*SD*	*Mdn*	*SD*	Statistic	Z-value
General Knowledge Task Pearson's Correlation Coefficient	.50	.19	.40	.19	641.00	-1.68
General Knowledge Task Discrimination Index	2.42	1.47	2.58	1.57	796.00	-.22
Emotion Recognition Task Pearson's Correlation Coefficient	.34	.22	.04	.25	554.50	-2.50*
Emotion Recognition Task Discrimination Index	1.28	1.15	.16	.96	466.00	-3.34**
CSSQ Task Pearson's Correlation Coefficient	.00	.36	.00	.24	677.00	-1.62
CSSQ Task Discrimination Index	8	3.98	8.88	4.65	729.50	-.85

*$p < .05$, **$p < .01$

Note: CSSQ = Child Social Situation Questionnaire

Hypothesis 3. It was expected that all children would be overconfident on difficult items but underconfident or better calibrated on easy items of the General Knowledge Task. Moreover, given the expectation that children in the clinical ADHD group will be less well calibrated overall (hypotheses 1 and 2), it was predicted that while both groups will be overconfidence on difficult items, the clinical ADHD group may demonstrate more overconfidence on difficult items and underconfidence on easy items relative to the TD group. A mixed between-within subjects analysis of variance was conducted to assess the impact of group (TD versus clinical ADHD) on participants' Bias Index Scores across Easy and Hard items on the GKT. There was no significant interaction between group and difficulty, Wilks' Lambda = .997, $F(1, 79) = .27$, $p = .61$, partial eta squared = .003. There was a substantial main effect for difficulty, Wilks' Lambda = .69, $F(1, 79) = 35.56$, $p \leq .001$, partial eta squared = .31, with participants providing underconfident Bias Index Scores on easy items and overconfident Bias Index Score on hard items (see Table 8). The main effect comparing the two groups was not significant, $F(1, 79) = .11$, $p = .75$, partial eta squared = .001, suggesting no difference between the TD and clinical ADHD group.

Table 8

Bias Index Differences on Hard vs. Easy Items GKT

	TD Group			Clinical ADHD Group		
Measure	n	*M*	*SD*	n	M	*SD*
General Knowledge Task Bias Index Hard Items	42	.11	.23	39	.13	.21
General Knowledge Task Bias Index Easy Items	42	-.01	.08	39	-.02	.09

Hypothesis 4. Based on the decision-making literature (Stanovich & West, 2000; Wolfe & Grosch, 1990), it was hypothesized that the Absolute Accuracy Index would be significantly negatively associated (lower absolute accuracy means better calibrated) with estimated intelligence and EFs. Spearman's Rank Order correlations were used to correlate cognitive abilities with performance calibration using the Absolute Index in order to determine if more accurate children also had higher cognitive abilities. These correlations were obtained for the

entire sample in Table 9. All of the cognitive ability measures were significantly intercorrelated (i.e. IQ composite z-score, Trailmaking difference score, and Stroop difference score). The IQ composite z-score was significantly negatively correlated with the GKT Absolute Accuracy Index; Trailmaking was significantly positively correlated with the GKT Absolute Accuracy Index and the ERT Absolute Accuracy Index; and the Stroop Interference score was significantly correlated with the GKT Absolute Accuracy Index. Therefore, the higher the cognitive abilities (higher IQ, lower Trailmaking, and lower Stroop Interference score), the better (lower) absolute accuracy on the GKT.

Table 9

Correlations between Absolute Accuracy and Cognitive Abilities for the Entire Sample

Measure	1.	2.	3.	4.	5.	6.	7.
1. General Knowledge Task Absolute Accuracy Index	1						
2. Emotion Recognition Task Absolute Accuracy Index	.18	1					
3. CSSQ Task Absolute Accuracy Index	.002	.19	1				
4. IQ Composite z-score	-.61**	-.13	-.04	1			
5.Trailmaking Part B minus Part A Time	.32**	.24*	.04	-.39**	1		
6. Stroop Interference Score	.55**	.05	-.10	-.30**	.36**	1	
7. BDEFS Summary Score	.14	.06	-.12	-.13	.09	.13	1

**p* < .05, ***p* < .01
Note: CSSQ = Child Social Situation Questionnaire, BDEFS = Barkley Deficits in Executive Functioning Scale

Hypothesis 5. It was predicted that participants in the clinical ADHD group would rate the performance calibration tasks as more difficult and effortful than the TD group and that the clinical ADHD group would have less of an association between perceived difficulty confidence judgments. Median parent ratings of their child's metacognitive skills and median ratings of cognitive effort across all tasks were compared across groups with Independent Samples Mann-Whitney U tests in Table 10. Median parent ratings of their child's metacognitive skills and median ratings of cognitive effort employed on the GKT, ERT, CSSQ, KBIT and Trailmaking significantly differed between groups. Thus, the TD group; had parents who rated their metacognitive skills as higher than those in the clinical ADHD group; and consistently reported more effort exerted on the GKT, ERT, CSSQ, KBIT and Trailmaking as compared to the clinical ADHD group.

Table 10

Group Differences in Parent Metacognition Rating and Task Difficulty and Effort Ratings

	TD Group		Clinical ADHD Group		Independent Samples Mann-Whitney U Test	
Measure	*Mdn*	*SD*	*Mdn*	*SD*	Statistic	Z-value
Parent Metacognition Rating	6	.88	5	1.37	444.00	-3.66**
General Knowledge Task Task Difficulty	1	.79	1	1.10	755.00	-.65
Emotion Recognition Task Task Difficulty	0	.77	0	.79	808.00	-.12
CSSQ Task Difficulty	2	1.01	2	1.31	799.50	-.19
KBIT Task Difficulty	2	1.07	2	1.07	716.00	-1.02
Trailmaking Task Difficulty	2	1.07	2	1.26	813.50	-.05
Stroop Task Difficulty	3	1.21	3	1.18	785.00	-.13
General Knowledge Task Cognitive Effort	5	.83	4	1.48	531.00	-2.92**
Emotion Recognition Task Cognitive Effort	5	1.08	4	1.71	591.50	-2.28*
CSSQ Cognitive Effort	5	.73	4	1.29	562.00	-2.72**
KBIT Cognitive Effort	5	.83	4	1.04	513.00	-3.08**
Trailmaking Cognitive Effort	5	.80	4	1.23	546.00	-2.92**
Stroop Cognitive Effort	5	.77	5	1.08	658.00	-1.58

*$p < .05$, **$p < .01$

Note: CSSQ = Child Social Situation Questionnaire, KBIT = Kaufman Brief Intelligence Test

In order to determine the association between perceived difficulty and confidence, a method similar to the discrimination indices for performance calibration was used. Task Difficulty and Effort were correlated using Spearman-Rank Order correlations, with the mean confidence rating for each of the General Knowledge Task, Emotion Recognition Task, and the Child Social Situation Questionnaire in order to determine whethr perceived difficulty is associated or calibrated with how confident one is in their answer. In other words, if a task is perceived to be difficult, then this should be reflected by a lower mean confidence rating for that task. Only the ERT task confidence was correlated with ERT task difficulty ($r_{\text{entire sample}} = -.33$, $p < .01$, $r_{\text{clinical sample}} = -.64$, $p < .01$). That is, the harder the ERT was perceived to be, the lower the mean confidence rating for this task for the entire sample and the clinical sample.

Hypothesis 6. It was hypothesized that clinical ADHD group would have worse performance calibration than the TD group. A mixed between-within subjects analysis of variance was conducted to assess the impact of group (TD versus clinical ADHD) on participants' Bias Index Scores across emotions on the ERT (sad, happy, angry, afraid, and neutral). There was no significant interaction between group and emotions, Wilks' Lambda = .91, F (4, 76) = 1.96, p = .11, partial eta squared = .09. There was a substantial main effect for emotion, Wilks' Lambda = .43, F (4, 76) = 25.46, $p \leq .001$, partial eta squared = .57, (see Table 11), such that sadness was the emotion that was rated most overconfident followed by angry, with happy afraid and neutral more accurately calibrated. The main effect comparing the two groups was significant, F (1, 79) = 4.03, p = .048, partial eta squared = .05, suggesting a small to moderate effect size in the difference between the TD and clinical ADHD group. Specifically, participants in the TD group were underconfident over most emotions while participants in the clinical ADHD group were overconfident overall.

Table 11

Bias Index Differences By Emotions on the ERT

	TD Group			Clinical ADHD Group		
Measure	n	*M*	*SD*	n	M	*SD*
ERT Bias Index Sad Items	42	.18	.22	39	.28	.24
ERT Bias Index Happy Items	42	-.02	.08	39	.003	.12
ERT Bias Index Angry Items	42	.06	.22	39	.05	.17
ERT Bias Index Afraid Items	42	-.04	.11	39	.02	.16
ERT Bias Index Neutral Items	42	-.01	.14	39	.06	.14

Note: ERT = Emotion Recognition Task

Hypothesis 7. It was expected that children in the clinical ADHD group would be better calibrated on the ambiguous negative social situations and less well calibrated on the positive social situations compared to the TD group. Independent Samples Mann-Whitney U Tests were used to compare Bias and Absolute Accuracy Index scores for Ambiguous Positive and Ambiguous Negative stories on the CSSQ between the two groups. Results are presented in Table 12. Significant median differences between groups were found for CSSQ Bias Index – Ambiguous Negative Stories such that participants in the TD group were slightly underconfident and participants in the clinical ADHD group were well calibrated. Significant median differences between groups were found for CSSQ Absolute Accuracy Index – Ambiguous Negative Stories such that, participants in the TD group had significantly higher Absolute Accuracy indicative of poorer performance calibration than the clinical ADHD group.

Table 12

Group Difference in Performance Calibration on Ambiguous Positive and Ambiguous Negative CSSQ Stories

	TD Group		Clinical ADHD Group		Independent Samples Mann-Whitney U Test	
Measure	*Mdn*	*SD*	*Mdn*	*SD*		Z-value
CSSQ Bias Index Ambiguous Positive Stories	-.08	.19	.00	.21	643.00	-1.67
CSSQ Bias Index Ambiguous Negative Stories	-.01	.16	.00	.13	578.50	-2.36*
CSSQ Absolute Accuracy Index Ambiguous Positive Stories	.05	.09	.01	.18	659.50	-1.52
CSSQ Absolute Accuracy Index Ambiguous Negative Stories	.03	.13	.00	.16	619.00	-1.97*

*$p < .05$

Note: CSSQ = Child Social Situation Questionnaire

Exploratory Analyses Based on Results

Predictors of performance calibration on the GKT, ERT, and CSSQ task.

Three simultaneous regressions were carried out in order to determine whether cognitive abilities and EFs were significant predictors of performance calibration the 1) GKT, 2) ERT, and 3) CSSQ Task. All independent variables (IQ and EFs) were entered simultaneously into each model.

The linear combination of IQ, Trailmaking Difference Score, and Stroop Difference Score significantly predicted GKT Absolute Accuracy (see Table 13). Controlling for Trailmaking and Stroop, IQ significantly predicted 15% of the variance in GKT Absolute Accuracy. Controlling for Stroop and IQ, Trailmaking was not a significant predictor of GKT

Absolute Accuracy. Controlling for IQ and Trailmaking, Stroop was a significant predictor or GKT Absolute Accuracy, accounting for 8% of the variance in the GKT Absolute Accuracy Index. The linear combination of IQ, Trailmaking Difference Score, and Stroop Difference Score did not significantly predict the ERT Absolute Accuracy Index, nor did they significantly predict the CSSQ Absolute Accuracy Index, $F(3,76) = .59$, $p = .62$, and $R^2 = .02$, $F(3,76) = .63$, $p = .60$ respectively.

Table 13

Simultaneous Regression Analysis of Intelligence and Executive Functions Predicting Performance Calibration on the GKT (Absolute Accuracy Index)

Variable	*B*	*SE(B)*	*t*-score	sr^2
IQ z-score Composite	-.02	.004	-4.62*	.15
Trailmaking Difference Score	.00	.00	1.43	.01
Stroop Difference Score	.001	.00	3.49*	.08

$R^2 = .48$, $F(3,76) = 23.19$, $p < .001$
*$p < .05$

Sex differences in performance and confidence on metacognition tasks.

There were no sex differences in accuracy or confidence on metacognition tasks as depicted in Table 14.

Table 14

Sex Differences in Performance and Confidence on Metacognition Tasks

	Males		Females		Independent Samples Mann-Whitney U Test	
Measure	*Mdn*	*SD*	*Mdn*	*SD*	*Statistic*	*Z-value*
General Knowledge Task Score /36	30.00	3.44	30.00	3.36	635.00	-.81
Emotion Recognition Task Score /40	34.00	2.60	34.50	2.21	614.50	-1.03
CSSQ Task Score /8	8.00	1.31	8.00	.63	548.00	-1.91
General Knowledge Task Mean Confidence Rating	8.56	1.13	8.53	.67	694.00	-.21
Emotion Recognition Task Mean Confidence Rating	9.38	.95	9.13	.95	548.00	-1.69
CSSQ Task Mean Confidence Rating	9.38	.84	9.38	.94	647.50	.68

Note: CSSQ = Child Social Situation Questionnaire

Discussion

The current study investigated performance calibration for general knowledge, emotion recognition, and recognition of the valence of social situations. See Table 15 for a summary of findings.

Table 15

Summary of Performance Calibration Findings

	GKT	ERT	CSSQ
Group Differences in Accuracy			
Group Differences in Confidence		X	X
Hypothesis 1: Group Differences in Absolute Accuracy - Absolute Accuracy Index			
Hypothesis 1: Group Differences in Absolute Accuracy - Bias Index		X	X
Hypothesis 2: Group Differences in Relative Accuracy - Pearson’s Correlation Coefficient		X	
Hypothesis 2: Group Differences in Relative Accuracy - Discrimination Index		X	
Hypothesis 3: Hard-Easy Effect	X		
Hypothesis 4: Absolute Accuracy Index Correlated with Cognitive Abilities	X	X*	
Hypothesis 5: Group Difference in Perceived Difficulty			
Hypothesis 5: Group Differences in Effort Expended	X	X	X
Hypothesis 5: Difficulty Calibrated to Confidence		X	
Hypothesis 6: ANOVA Main Effects		X	
Hypothesis 7: Group Differences in Ambiguous Positive Social Situations -Absolute Accuracy Index			
Hypothesis 7: Group Differences in Ambiguous Positive Social Situations - Bias Index			
Hypothesis 7: Group Differences in Ambiguous Negative Social Situations - Absolute Accuracy Index			X
Hypothesis 7: Group Differences in Ambiguous Negative Social Situations – Bias Index			X

Note: X indicates a significant finding.
* only for Trailmaking.

When performance calibration was examined in terms of the accuracy of items completed and mean confidence ratings, the clinical ADHD group had significantly higher confidence scores on the ERT and the CSSQ. No sex differences were found in accuracy on items or confidence levels on the GKT, ERT, nor the CSSQ which is consistent with other research that has found that females can be equally as overconfident as males (Allwood, Johnsson, & Granhag, 2005; Lundeberg & Mohan, 2010). The clinical ADHD group had significantly higher symptoms and impairment for ADHD, ODD, and CD in the past year than the TD group, which is not surprising given that some reports estimate 40% to 90% of children with ADHD will also have comorbid diagnoses in either ODD or CD (Tannock, 1998). In terms of cognitive abilities, the clinical ADHD group had significantly lower nonverbal reasoning and took significantly longer to shift sets which is in line with other research confirming cognitive and executive dysfunction in ADHD (Faraone et al., 2015; Seidman, 2006).

Following, there will first be a discussion of both absolute and relative accuracy across groups and domains alongside recommendations for future use of indices. The impact of item difficulty and cognitive abilities on performance calibration is then described and examined in the context of current research. Next, group differences in identification of emotions and interpretations of social situations are compared across absolute accuracy indices along with parent ratings of child performance calibration, and child self-reports of difficulty and effort. Finally application for interventions and assessment, limitations, implications and future directions are discussed.

Absolute Accuracy across Groups and Domains

Recall that absolute accuracy is measured by the Absolute Accuracy and Bias indices. As hypothesized, children in the clinical ADHD group were significantly overconfident on the ERT while the TD group was less overconfident. The clinical ADHD group was well-calibrated (no under/overconfidence) on the CSSQ compared to the TD group who were underconfident. It should be noted here, as is discussed later, that since the CSSQ is composed of four *ambiguous negative* and four *ambiguous positive* situations, and it may be more adaptive here to display underconfidence since the nature of the situations are ambiguous, some modesty or caution should be reflected in the confidence ratings. Therefore, the median Bias Index score of zero for the clinical ADHD group does not do a good job of reflecting the most adaptive response since high confidence was placed on negative ambiguous situations. These results replicate findings that children with ADHD have difficulties with self-awareness and reflection (Barkley, 1996) that often results in a PIB or the perception that one is doing better than other raters report (Evangelista et al., 2008; Gerdes et al., 2003; Hoza et al., 2004; Scholtens et al., 2012; Owens & Hoza, 2003). Additionally, high symptom rates of ADHD have been associated with overestimation/overconfidence in social and academic domains (Fefer et al., 2015). Interestingly, only the ERT and CSSQ Bias Index scores were different amongst groups. There were no significant differences on the GKT likely because children in both groups received similar high scores on these tasks and also had similar mean confidence ratings. It is possible that there was not a large enough range of difficulty of items on the GKT to obtain group differences in confidence ratings since the median score on the task was 30/36 indicating a potential ceiling effect. Nonetheless, these findings suggest no group differences when easy items are administered.

It is noteworthy that while there were differences in overconfidence, there were no group differences in the Absolute Accuracy Index across the ERT and CSSQ. The Absolute Accuracy Index is a measure of item specific discrepancy and produces an absolute value. Higher Absolute

Accuracy scores are indicative of poorer performance calibration. Conversely, the Bias Index is a measure of under/overconfidence and it is here that differences were obtained indicating that the clinical ADHD was more overconfident on the ERT and CSSQ tasks. In other words, in absolute terms, the groups did not differ on the ERT and CSSQ, but in terms of under/overconfidence, the clinical ADHD group was significantly overconfident. This overconfidence bias towards the heavily social tasks ties into social information processing (SIP) as a factor that can influence how one interprets and attributes social cues, thereby influencing goal selection. Based on the current study, children with ADHD show overconfidence bias in both of the interpretation of social cues (i.e. facial expressions) and in the interpretation of social situations. Additionally, the Absolute Accuracy scores across various domains significantly differed such that the CSSQ was the best calibrated task and the GKT and ERT were less well calibrated overall.

Relative Accuracy across Groups and Domains

When performance calibration in terms of relative accuracy was examined, there were group differences in both the Pearson’s Correlation Coefficient and Discrimination Index but only for the ERT, partially supporting the hypothesis that groups would differ in relative accuracy. The clinical ADHD group had significantly lower ERT Pearson’s Correlation Coefficient scores, indicating that there was less of an association between accuracy ratings and confidence judgments. The TD group had a significantly higher ERT Pearson’s Correlation Coefficient which indicates that this group was better able to adjust their confidence ratings in relation to how accurate they thought they were on a given item. In other words, there was a stronger association/correlation between accuracy and confidence for the TD group as compared to the clinical ADHD group suggesting that for the ADHD group, confidence ratings were less strongly associated with accuracy for recognizing emotions on faces.

The clinical ADHD group had significantly lower ERT Discrimination Index scores, indicating that this group was less able to monitor and adjust confidence across correct and incorrect items. This deficit in self-monitoring and adjusting may further impact SIP during the interpretation of self and other attributions and social cues and exacerbate negative biases. The TD group had significantly higher ERT Discrimination Index scores which means that this group was significantly more confident than the clinical ADHD group for correct items versus incorrect items (i.e. they did not as misplace high confidence scores with incorrect items). This finding supports the hypothesis that the TD group were better than the clinical ADHD group at adjusting their confidence ratings according to their accuracy. However, this hypothesis was only partially supported, as this difference was only found for the ERT and not the GKT nor the CSSQ. One reason why group differences were not found on the GKT and the CSSQ may again be due to the range of difficulty of the items. There seems to have been ceiling effects, suggesting that the items were easy for the children. As discussed previously, easier items tend to be better calibrated than difficult items (Juslin et al., 2000; Lichtenstein & Fischhoff, 1977) and therefore lacking a range of difficulty could have made it difficult to find group differences in confidence since both groups were likely well calibrated on easy items.

There were consistent group differences in performance calibration on the ERT specifically, and this assertion is further supported by the finding that overall, the harder the ERT was perceived to be, the lower the confidence rating. The correlation between confidence and difficulty rating was negative and significant implying that the more difficult the ERT was perceived to be, the lower the confidence rating.

Evaluation of Performance Calibration Indices

Given concerns about the reliability of performance calibration measures in terms of number of judgments made and delayed vs. immediate judgments, (Thompson, 1999; Veenman et al., 2006), the current study addressed these concerns with many items per domain, by using both concurrent (item by item confidence judgments) and retroactive judgments (perceived difficulty and effort), as well as by employing absolute and relative accuracy indices. All of the measures were found to be reliable. In terms of evaluating the usefulness of each of the four indices used, it is important to remember that each index is measuring a different construct.

Absolute accuracy, the *difference* between confidence judgments and performance on a task, was measured in the current study by the Absolute Accuracy Index and the Bias Index. Indeed, important differences emerged between the two measures. With the Absolute Accuracy Index measuring performance calibration in absolute terms, few group differences emerged. The Absolute Accuracy Index does not provide information to determine the direction of the miscalibration, and therefore both under and overconfidence are treated equally in absolute terms. Using the Absolute Accuracy Index, no group difference emerged across general knowledge, emotion recognition, and social situation tasks. When the Bias Index was examined, group differences emerged.

A lack of group differences with the Absolute Accuracy Index and finding group differences with the Bias Index is an important indicator of the utility of each of these indices. Since the current study aimed to investigate not only *whether* the groups differed in their performance calibration, but also *how* the groups differed, the Bias Index provided more valuable information about how the groups calibrated their performance. For instance, while no group differences emerged in the ERT Absolute Accuracy Index nor the CSSQ Absolute

Accuracy Index, the Bias Indices for these measures told a different, more detailed story. In fact, on the ERT Bias Index, the TD group was only somewhat overconfident on average, while the clinical ADHD group was significantly more overconfident. We can see the utility of the Bias Index most clearly with regard to the CSSQ, as the TD group was found to significantly differ from the clinical ADHD group, with the TD group underconfident, and the clinical ADHD overconfident. The difference in under/overconfidence is overlooked with the Absolute Index as only absolute differences between performance and confidence ratings are considered. As such, the Bias Index is preferable for use when one is interested in not only *whether* groups differ in performance calibration, but also *how* groups differ in performance calibration.

Relative accuracy indices are helpful in the examination of the *association* between confidence judgments and performance, where good performance calibration entails confidence ratings be matched with performance on a task. The current study aimed to assess whether the groups differed in their ability to adjust their confidence rating to match their performance. This was achieved with the Pearson's Correlation Coefficient, a correlational measure examining the association between confidence and performance; and the Discrimination Index, which assessed discrimination in confidence ratings for correct and incorrect items. Group differences emerged on the ERT Relative and Discrimination Indices only, suggesting that the TD group was better at matching confidence to performance and distinguishing confidence ratings for correct and incorrect items. The information gleaned from relative indices is different and informative for interpreting how groups differ in their performance calibration. While relative accuracy indices were not the focus of the current study, the results suggest that these measures may be particularly useful when the goal is to assess the ability to discriminate between correct and incorrect items with confidence ratings.

The Hard-Easy Effect

Contrary to the hypothesis, there were no interaction effects for the difficulty of items (hard and easy) and group (clinical ADHD group and TD group) for the GKT. This finding was surprising given that it was both expected and confirmed in the first two hypotheses that children with ADHD were more misacalibrated than children without ADHD, and also this finding should be more pronounced with difficult items given that item difficulty has been found to increase overconfidence (Klayman & Soll, 1999). There were suggested ceiling effects for the GKT and as such, only 12 items were identified as difficult and this may not have been a sufficient number of difficult items to compare across groups.

There was a significant main effect for difficulty such that easy items on the GKT elicited underconfidence and hard items elicited overconfidence. This is consistent with other research examining the effect of item difficulty on confidence ratings. The *hard-easy* effect predicts that when accuracy on a set of items is high, confidence is low (underconfidence) and when accuracy is low, confidence is high (overconfidence) (Juslin et al., 2000; Lichtenstein & Fischhoff, 1977). Part of the rationale for this prediction is that people who are lacking skills or knowledge in a particular domain are not only incorrect in their performance but because they lack information they are also impaired in their ability to recognize their errors as errors and therefore have poor performance calibration (Kruger & Dunning, 1999).

Performance Calibration and Cognitive Abilities

The hypothesis that better performance calibration (lower Absolute Accuracy Index) would be associated with better cognitive abilities was only partly supported by the data. As expected, all cognitive abilities were significantly intercorrelated. The association between the GKT Absolute Accuracy Index and the IQ composite was negative, moderate, and significant suggesting that children with higher intelligence were more accurate in their assessment of their performance for GKT type of questions. The association between the GKT Absolute Accuracy Index and the Trailmaking difference score was positive, weak, and significant suggesting that children with better set-shifting (lower time difference scores) were also more accurate in their assessment of their performance for GKT type of questions. The association between the GKT Absolute Accuracy Index and the Stroop Interference score was also negative, moderate, and significant suggesting that children with a better ability to stop a prepotent response (inhibition) also were more accurate in their assessment of their performance for GKT type of questions. Moreover, cognitive abilities significantly predicted the Absolute Accuracy Index scores for general knowledge items. Overall, better cognitive abilities were associated with and predictive of better performance calibration which is consistent with research finding a negative association between cognitive abilities and overconfidence (Parker & Fischhoff, 2005; Stanovich & West, 2000; Wolfe & Grosch, 1990). However, this association does not hold across all domains.

Performance calibration in terms of the Absolute Accuracy Index on the ERT was weakly, positive, and significantly associated with the Trailmaking score, or set-shifting only; while the Absolute Accuracy Index on the CSSQ was not significantly associated with any of the cognitive ability measures. These findings suggest that performance calibration in SIP domains are not necessarily associated with cognitive abilities for the current sample. Other research has also suggested that with age, performance calibration tends to rely less on general cognitive abilities including EFs and that models of EF are not able to provide satisfying accounts of performance calibration (del Missier, Mäntylä, & Bruine de Bruin, 2012). Instead, performance on SIP type tasks may rely more heavily on the ability to detect emotional cues and interpret social situations than on cognitive abilities.

Performance Calibration and Emotion Recognition

While there was no interaction between group (clinical ADHD group and TD group) and emotions (happy, angry, afraid, neutral, and sad), there was a substantial main effect for emotions and main effect for group, thus, supporting the hypothesis that children with ADHD would have worse performance calibration. Findings indicated that the most miscalibrated emotion was sad, followed by angry, afraid, neutral and happy. These results provide further evidence for the finding that sadness and anger are the least accurately recognized facial emotions out of happiness, fear, anger, sadness, and disgust in 4 to 16 year-olds (Montirosso et al., 2010).

There was also a main effect of group, with the TD group having more accurate or underconfident biased ratings and the clinical ADHD group being more overconfident overall. Even though the overall Bias Index score for the TD group was slightly overconfident, this group was underconfident on 3/5 of the emotions whereas the clinical ADHD group was overconfident on all emotions. The current study did not find the clinical ADHD group to be less accurate at recognizing facial emotions, contrary to findings that show children with ADHD to be less accurate in their recognition of facial emotions than TD children (Cadesky et al., 2000; Shapiro, Hughes, August, & Bloomquist, 1993). One reason for this may be ceiling effects, in that both groups did well on this task and therefore the range of emotions on the faces may not have been difficult enough or not complex enough to elicit a difference in the interpretation of facial cues. The current study adds to the literature in showing that while children with ADHD may or may not necessarily be impaired in their recognition of facial emotions, they often overestimate their abilities in doing so. This may have implications for SIP since in order to correct or decide on how to behave, one must process errors as such and correct behaviour accordingly. This finding has important implications in terms of intervention and strategies to help children with ADHD acknowledge when they have misinterpreted an emotional expression and how they respond to social situations.

Performance Calibration and Interpretation of Social Situations

The results supported the hypothesis that children in the clinical ADHD group are better calibrated on the ambiguous negative social situations but did not support the hypothesis that they would less well calibrated on the positive social situations compared to the TD group. As noted earlier, interpretation of these findings is reliant on the nature of the situations presented in the CSSQ. The CSSQ is composed of four *ambiguous negative* and four *ambiguous positive* situations. On this measure, it may be adaptive to have lower confidence, since the nature of the situations are ambiguous, some modesty or caution should be reflected in the confidence ratings.

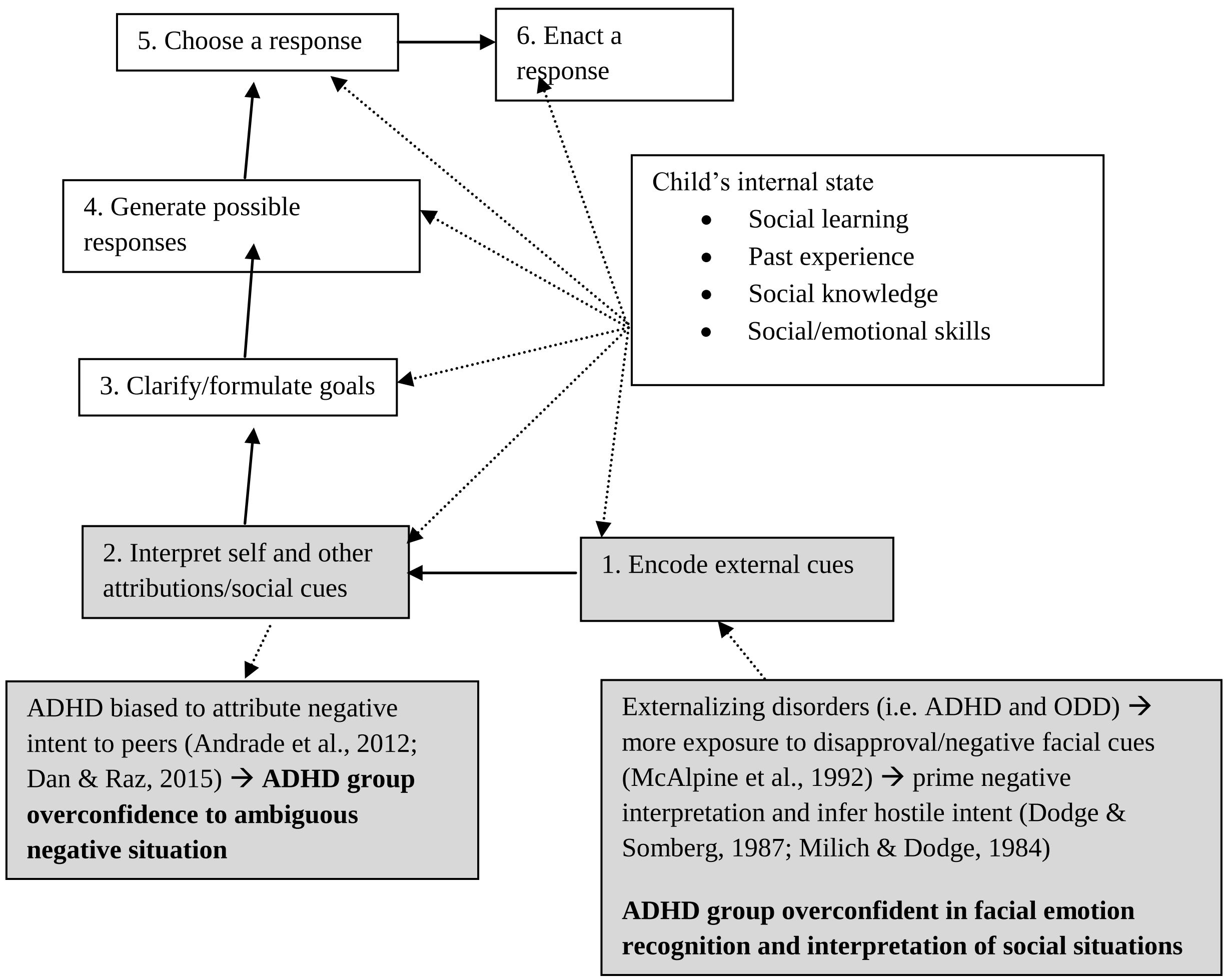
First, it should be noted that there was a ceiling effect for performance on this task in that most children correctly answered whether the situation was positive or negative. There were no significant group differences in performance calibration on the absolute accuracy measures (Absolute Accuracy Index and Bias Index) for ambiguous positive situations. This finding is consistent with the ceiling effect, in that, children in both groups were accurate and equally as confident. There were, however, significant group differences for the ambiguous negative situations. Here, the children in the clinical ADHD group were perfectly calibrated (absolutely and not under/overconfident) while the TD group was underconfident on the ambiguous negative situations. Given that the situations are ambiguous, being well calibrated may in fact mean modesty in confidence ratings or underconfidence, since the situation is ambiguous the confidence should reflect this and be slightly lower. While children in the clinical ADHD group were perfectly calibrated on the negative ambiguous situations, it must be noted that they were certain about the valence of a situation that was in itself designed to be somewhat ambiguous. This finding is in line with the supposition that deficits in SIP are fundamental to the maladaptive interpersonal functioning that many children with ADHD experience. Deficits encoding external cues and interpreting others attributions are importantly at play when deciding whether a situation is positive or negative. The nature of the ambiguous situations calls on the individuals' social learning, past experience and social knowledge, which may be biased to attribute negative intent to peers (Andrade et al., 2012), in turn inflating confidence for ambiguous negative situations. When considered in this light, one could argue that it was the TD group that was better calibrated since they took this ambiguity into account in their confidence ratings and as such resulted in overall underconfidence for this task. So then why did children in the ADHD group calibrate their performance on the *negative ambiguous* situations so perfectly when they maybe should have been somewhat underconfident?

As demonstrated, SIP in children with ADHD has been found to be impaired in the detection and encoding of external cues (Andrade et al., 2012) and may be negatively biased (Dan & Raz, 2015). The impact of inaccurate detection of facial cues and how children represent/attribute a social situation for SIP is unclear (Orobio de Castro, 2004). Research on the recognition of facial emotions suggests that one's ability to recognize emotions may be due in part to exposure to that emotion (McAlpine et al., 1992) and therefore children with externalizing disorders such as ADHD and ODD may have more exposure to disapproval and negative facial expressions. This exposure may then prime these children toward negative interpretation as boys with ADHD have been found to infer hostile intent in others' actions (i.e. negatively primed/biased) when emotionally aroused (Dodge & Somberg, 1987; Milich & Dodge, 1984).

Moreover, other studies found that children with ADHD may be more accurate at the recognition of anger or negative affect than TD controls (Cadesky et al., 2000). Since the situations were negative and ambiguous in this study, this priming may have caused the participants in the clinical ADHD group to be correct and confident, without the ambiguity of the situation impacting their confidence ratings. The current study supports the notion that children with ADHD have deficits in their SIP, specifically with regards to encoding external cues and interpreting other's attributions. This study extends SIP theory by suggesting that children with ADHD may also be impaired and biased at interpreting their own internal cues for the accuracy of their judgments in the interpretation of external cues and social situations. See Figure 2 for a visual depiction of how the current study supports, extends, and impacts SIP theory.

Figure 2.

Modified Social Information Processing Model (Crick & Dodge, 1994)



Parent Ratings of Performance Calibration and Self-Reports of Difficulty and Effort

As expected, parents of children in the clinical ADHD group rated their children's metacognitive performance calibration skills as significantly lower than parents of the TD group. Given that models of ADHD encompass difficulties in self-awareness and reflection as core components of the disorder (Barkley, 1996) including deficits in strategy use (Hervey et al., 2004), and that avoidance of effortful tasks is a symptom criteria for ADHD, it was expected that children with ADHD should perceive tasks to be more difficult and effortful than controls.

Interestingly, while there were no group differences in perceived difficulty across the GKT, ERT, CSSQ, KBIT, Trailmaking, and Stroop, the TD group consistently reported more effort exerted on the GKT, ERT, CSSQ, KBIT and Trailmaking as compared to the clinical ADHD group. This finding is both novel and remarkable as it points to a specific deficit in motivation for children with ADHD. If both groups perceive the task to be equally as difficult, then why do children with ADHD report that they do not exert the same amount of effort? One answer could be that these children do not know the appropriate amount of effort to exert given a particular level of difficulty. The explanation lends itself to the notion that children with ADHD are miscalibrated in many ways, one way being the calibration of effort to difficulty. This finding has important implications for interventions aiming to help children improve their metacognition and performance calibration as it provides evidence for a specific deficit, namely, matching effort expended to perceptions of difficulty.

Performance Calibration as a Point for Intervention

Cognitive-behavioural approaches to the treatment of ADHD have involved self-reinforcement, problem-solving, self-instruction, and self-redirection to cope with errors (Toplak, Connors, Shuster, Knezevic, & Parks, 2008). In such approaches, it is believed that self-control will improve by enhancing cognitive or metacognitive skills that are thought to underlie impulsivity and goal-directed behaviour. Given that the development of metacognition is reliant on metastrategic knowing or the ability to select useful strategies and eliminate unsuccessful strategies (Kuhn, 2014), having poor performance calibration can be an impediment to metacognitive development and therefore lends itself well to a focus for intervention strategies. Confidence in what we know is an important part of metacognition given that in order to know where to focus effort in learning new information, one must be able to accurately judge one’s knowledge base (Lundeberg & Mohan, 2010). In the current study, performance calibration in children with ADHD has been shown to be impaired in the recognition of facial affect and interpretation of social situations. Given these findings, immediate confidence ratings of correctness or the prompting to reflect on performance may provide an important means for differentiating between groups as well as an avenue for intervention (Parker & Fischhoff, 2005).

Directing children to their internal cues and self-reflection on the accuracy of their performance in different domains can also provide an avenue for teaching children how to make better decisions (Weller et al., 2012). Interventions that target and teach decision-making to children have the potential to improve social, financial and health outcomes later in life (Baron & Brown, 1991). For instance, strategies that encourage the consideration of the opposite of the decision that has been made (Larrick, 2004) or requiring that one justify his/her decision to another or to take an outsider's perspective (Lerner & Tetlock, 1999) may be useful in reducing overconfidence. Some recommendations for improving overconfidence with investors and executives include keeping score about the absolute accuracy of past decisions, consulting with others, and using baserates (Mauboussin & Callahan, 2015). Additionally, performance calibration for judgments of the probability of an event occurring has been found to improve when comprehensive feedback on performance is provided (Lichtenstein & Fischhoff, 1980). Taken together, these findings offer a plethora of strategies that may be employed in treatment of children with impaired performance calibration, such as children with ADHD. The effectiveness of training and treatment for self-monitoring or performance calibration in children with ADHD has been found socially, academically, and for behavioural outcomes.

In a study of boys aged 8-13 with ADHD, a cognitive-behavioural reinforced self-evaluation treatment was evaluated (Hinshaw, Henker, & Whalen, 1984). The treatment involved explicit training in self-monitoring and performance calibration with skills and concepts and was found to be effective at improving social interactions. In a more recent study, Joseph and Eveleigh (2011) examined the effect of specific strategies such as self-monitoring (performance calibration) on reading performance for students with ADHD. These authors found overall positive effects of self-monitoring on accuracy of words read correctly and the number of comprehension questions answered. Specifically, for students with ADHD, there was improved reading performance and on-task behaviour. Moreover, in a study examining college students with ADHD, participants received treatment in either skills instruction, goal setting, and self-monitoring instruction *or* only study skills and goal setting (Scheithauer & Kelley, 2014). Participants who received the treatment including self-monitoring instructions demonstrated improvement not only in their academic behaviour, GPAs, and goal attainment, but also had improvement in their ADHD symptoms while the group who did not receive the self-monitoring instruction did not show such improvement.

One paper provided a review of studies that targeted specific self-monitoring strategies such as self-recording checklists, prompts/cues to self-record, accuracy checking, training and reinforcement for a variety of disabilities, including learning disorders, emotional/behavioural disorders and intellectual disabilities (Webber, Scheuermann, McCall & Coleman, 1993). With regard to behavioural outcomes for interventions employing self-monitoring, the majority of studies found a positive effect on behavioural outcomes (i.e. on-task behaviour, task completion, and emotional outbursts). Another review found similar results in that self-monitoring was found to decrease off-task behaviour, disruptions, negative social interactions, improve on-task behaviour, completion of work, and positive social interactions (Bruhn, McDaniel, & Kreigh, 2015).

Limitations, Implications, and Future Directions

Ceiling effects.

The average scores on the GKT, ERT, and CSSQ were quite high. The difficulty of the items was intentionally set to be relatively easy as harder items tend to inflate confidence unnecessarily. There were no group differences in performance calibration on the GKT and few differences on the CSSQ which seem to be a result of ceiling effects for both groups. In the future, studies should improve the range of the difficulty of items presented in order to better examine how children with and without ADHD calibrate their performance across domains and difficulty of items.

Fatigue effects.

Participants who completed the performance calibration tasks at the end of the battery (i.e. Version B) had lower ERT and CSSQ accuracy scores than those who completed this task at the beginning of the battery, suggesting fatigue effects. In the current study, this effect was balanced across the entire sample and so it is not thought to impact group differences. Fatigue effects have been found to increase with increasing time-on-task but may also be impacted by personality, interest, and motivation (Ackerman & Kanfer, 2009). Fatigue effects have also been found to compromise executive control (Van der Lindin, Frese, & Meijman, 2003) which has been closely linked to metacognitive abilities (Dougherty & Sprenger, 2006; Weller et al., 2012). Therefore, future studies should have participants complete performance calibration measures at the beginning of a battery and try to limit the time on task or offer breaks in between tasks to limit the effects of fatigue on performance.

Sample size.

The sample size in the current study provided sufficient power to test the hypotheses, however, the sample size did not permit more elaborate analyses or comparisons of within group differences. Given that this study is the first of its kind to examine performance calibration in children with and without ADHD across these various domains, it provides compelling evidence for future studies to examine these domains and group differences in larger samples.

Confounds.

It could be argued that the paradigms and measures for performance calibration used in the current study conflate confidence for judgments because they only ask about the confidence in the accuracy of one’s answer and not the inaccuracy of it. Individual biases in modesty or conviction about accuracy may differ dependent on whether one believes they answered correctly or incorrectly. For example, someone may be very unsure if they were correct, and conversely they may be confident that they were incorrect. Future studies could address this limitation by asking whether respondents think they answered the question correctly and then examine their confidence in the correctness and the incorrectness of their answers.

Since there was a high number of children with comorbid ODD and a few with CD, it cannot be determined whether differences in performance calibration between the clinical ADHD group and the TD group are due to ADHD or the comorbidities. Unfortunately, given the sample size of comorbidities, it was not possible to parse apart these effects. While it is ecologically valid to include children with ADHD who have comorbidities, future studies should attempt to include more subgroups of ADHD with and without comorbidities to better address core impairments in performance calibration.

Social desirability was not measured in the current study and therefore, if for example, the TD group was more motivated or inclined to try harder or exert more effort (or the clinical ADHD group exerted less effort), this study would not be able to determine whether this factor was impacting any of the results. Future studies could include measures of social desirability to better determine whether social desirability impacts performance calibration.

Implications and Future Directions

The present study was novel in that it examined performance calibration in a clinical sample of children with ADHD and TD children, and assessed this construct across a variety of domains. Evidence was found to support that notion that children with ADHD are impaired in their self-monitoring and awareness in SIP tasks using a cognitive science paradigm of performance calibration. Recommendations for the use of different performance calibration indices have been made for future use, dependent on the research questions posed. These findings can also be incorporated into the treatment of SIP deficits in children with ADHD, and provide a basis for further research into where these children's SIP break down and lead to negative consequences. Furthermore, since group differences in performance calibration exist across these domains, future research could focus on the clinical utility of these measures to predict group membership. Finally, this study adds to the literature on performance calibration and cognitive abilities but also suggests that more research is needed in order to delineate the effects of cognitive abilities and EFs on performance calibration.

Conclusion

The current study investigated performance calibration indices for children with and without ADHD in the domains of general knowledge, emotion recognition, and recognition of the valence of social situations. Parents of children in the clinical ADHD group rated their children's performance calibration skills significantly lower than parents in the TD group. Children in the clinical ADHD group rated their effort expenditure as significantly lower on the GKT, ERT, CSSQ, KBIT, and Trailmaking than the TD group even though there were no differences in the perception of difficulty on these tasks. There were no group differences in general knowledge across performance calibration measures, although the *easy-hard effect* was demonstrated, such that harder items elicited overconfidence and easier items elicited underconfidence across groups. The only domain significantly associated with cognitive abilities was general knowledge such that higher cognitive abilities predicted better performance calibration.

Group differences in emotion recognition on the Bias Index scores were found, with the ADHD group more overconfident than the TD group. There were significant differences on the relative accuracy indices between groups on ERT Pearson's Correlation Coefficient and ERT Discrimination Index. Also, there was a significant negative correlation between confidence ratings and difficulty ratings overall on the ERT suggesting that for this domain, as confidence ratings went up, perceived task difficulty decreased. The TD group was more accurate or underconfident when recognizing emotions and the clinical ADHD group was overconfident across emotions. Additionally, a main effect of emotion indicated that the most miscalibrated emotion was sad, followed by angry, afraid, neutral and happy across groups. Regardless of impairment in the recognition of emotions, the clinical ADHD group was overconfident indicating that their metacognitive processes were inferring more confidence than they were accurate, which has important implications for their SIP.

On the CSSQ, group differences in the Bias Index were found, with the clinical ADHD group well calibrated and the TD group underconfident. The clinical ADHD group calibrated their performance on the *negative ambiguous* situation perfectly when they should have been somewhat underconfident, like the TD group. Performance calibration in SIP including emotion recognition on faces and interpretation of situations was not associated with the cognitive abilities measured in the current study. Additionally, the use of the Bias Index was recommended for future examinations of group differences in performance calibration across domains and an investigation of the clinical utility and predictive power of these differences was proposed. Thus, this is the first study using a cognitive science paradigm to assess performance calibration and find overconfidence in facial emotion recognition and interpretation of social situations for children with ADHD compared to a TD group.

References

Achenbach, T. M. (2011). Commentary: Definitely more than measurement error: But how should we understand and deal with informant discrepancies? *Journal of Clinical Child & Adolescent Psychology, 40*(1), 80-86.

Achenbach, T.M. (2013). *DSM-Oriented Guide for the Achenbach System of Empirically Based Assessment (ASEBA)*. Burlington, VT: University of Vermont Research Center for Children, Youth, and Families.

Achenbach, T. M., & Rescorla, L. (2001). ASEBA school-age forms & profiles.

Ackerman, P. L., & Kanfer, R. (2009). Test length and cognitive fatigue: An empirical examination of effects on performance and test-taker reactions. *Journal of Experimental Psychology: Applied*, *15*(2), 163.

Allwood, C. M., Jonsson, A. C., & Granhag, P. A. (2005). The effects of source and type of feedback on child witnesses' metamemory accuracy. *Applied Cognitive Psychology*, *19*(3), 331–344. http://doi.org/10.1002/acp.1071

Andrade, B. F., Waschbusch, D. A., Doucet, A., King, S., MacKinnon, M., McGrath, P. J., Stewart, S. H., & Corkum, P. (2012). *Journal of Attention Disorders, 16*(6), 491-504.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.

Balogh, L., & Czobor, P. (2014). Post-Error Slowing in Patients With ADHD: A Meta-Analysis. *Journal of Attention Disorders*. doi:10.1177/1087054714528043

Barell, J. (1992). Like an incredibly hard algebra problem: Teaching for metacognition. In A. L. Costa, J. A. Bellanca, & R. Fogarty (Eds.), *If minds matter: A foreword to the future* (Vol. 1). Palatine, IL: IRI/Skylight Publishing, Inc.

Barkley, R. A. (2012). *Barkley Deficits in Executive Functioning Scale-Children and Adolescents (BDEFS-CA)*. Guilford Press.

Barkley R. A. (1996). Linkages between attention and executive functions. In G. R. Lyon & N. A. Krasnegor (Eds.), *Attention, memory, and executive function*. Baltimore, MD: Brooks.

Barkley, R. A. (2006). *Attention-deficit hyperactivity disorder: A handbook for diagnosis and treatment*. New York, NY: Guilford Press.

Baron, J., & Brown, R. V. (1991). *Teaching decision making to adolescents*. Hillsdale, NJ: Erlbaum.

Barrera, M. E., & Maurer, D. (1981). The perception of facial expressions by the three-month-old. *Child Development*, *52*(1), 203–206. doi:10.2307/1129231

Boakes, J., Chapman, E., Houghton, S., & West, J. (2008). Facial affect interpretation in boys with attention deficit/hyperactivity disorder. *Child Neuropsychology, 14*, 82-96.

Bruhn, A., McDaniel, S., & Kreigh, C. (2015). Self-Monitoring Interventions for Students with Behavior Problems: A Systematic Review of Current Research. *Behavioral Disorders*, *40*(2), 102.

Bruine de Bruin, W., Parker, A. M., & Fischhoff, B. (2007). Individual differences in adult decision-making competence. *Journal of Personality and Social Psychology, 92*, 938–956.

Canadian Attention Deficit Hyperactivity Disorder Resource Alliance (CADDRA): Canadian ADHD Practice Guidelines, Third Edition, Toronto ON; CADDRA, 2011.

Cadesky, E. B., Mota, V. L., & Schachar, R. J. (2000). Beyond words: how do children with ADHD and/or conduct problems process nonverbal information about affect? *Journal of the American Academy of Child and Adolescent Psychiatry*, *39*(9), 1160–7. doi:10.1097/00004583-200009000-00016

Chan, T., & Martinussen, R. (2015). Positive illusions? The accuracy of academic self-appraisals in adolescents with ADHD. *Journal of Pediatric Psychology,* 1-11. doi: 10.1093/jpepsy/jsv116.

Chronaki, G., Hadwin, J. a, Garner, M., Maurage, P., & Sonuga-Barke, E. J. S. (2015). The development of emotion recognition from facial expressions and non-linguistic vocalizations during childhood. *British Journal of Developmental Psychology*, *33*, 218–236. http://doi.org/10.1111/bjdp.12075

Collin, L., Bindra, J., Raju, M., Gillberg, C., & Minnis, H. (2013). Facial emotion recognition in child psychiatry: a systematic review. *Research in Developmental Disabilities*, *34*(5), 1505-1520.

Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Bulletin*, *115,* 74-101. doi:10.1037/0033-2909.115.1.74

Dan, O., & Raz, S. (2015). Response Patterns to Emotional Faces Among Adolescents

Diagnosed With ADHD. *Journal of Attention Disorders*. http://doi.org/10.1177/1087054715606215

De Castro, B. O. (2004). The development of social information processing and aggressive behaviour: Current issues. *European Journal of Developmental Psychology, 1*(1), 87-102.

del Missier, F., Mäntylä, T., & Bruine de Bruin, W. (2012). Decision making competence, executive functioning, and general cognitive abilities. *Journal of Behavioral Decision Making, 25*, 331–351. doi: 10.1002/bdm.731

Denham, S. A., Way, E., Kalb, S. C., Warren-Khot, H. K.,&Bassett, H. H. (2013). Preschoolers' social information processing and early school success: The challenging situations task. *British Journal of Developmental Psychology, 31*, 180–197. doi:10.1111/j.2044-835X.2012.02085

Desoete, A., & Roeyers, H. (2006). Metacognitive macroevaluations in mathematical problem solving. *Learning and Instruction*, *16*(1), 12–25. doi:10.1016/j.learninstruc.2005.12.003

Dodge, K. A., & Somberg, D. R. (1987). Hostile attributional biases among aggressive boys are exacerbated under conditions of threats to the self. *Child Development, 58*, 213–234.

Dougherty, M. R., & Sprenger, A. (2006). The influence of improper sets of information on judgment: how irrelevant information can bias judged probability. *Journal of Experimental Psychology: General*, *135*(2), 262.

Draeger, S., Prior, M., & Sanson, A. (1986). Visual and auditory attention performance in hyperactive children: Competence or compliance. *Journal of Abnormal Child Psychology*, *14*(3), 411-424.

Dunlosky, J., Rawson, K. A., & Middleton, E. L. (2005). What constrains the accuracy of metacomprehension judgments? Testing the transfer-appropriate-monitoring and accessibility hypothesis. *Journal of Memory and Language, 52*, 551–565. doi:10.1016/j.jml.2005.01.011.

Edwards, R., Manstead, A. S. R., MacDonald, C. J., 1984. The relationship between children's sociometric status and ability to recognise facial expressions of emotion. *Eur. J. Social Psychol. 14*, 235–238.

Egger, H. L., Pine, D. S., Nelson, E., Leibenluft, E., Ernst, M., Towbin, K. E., & Angold, A. (2011). The NIMH Child Emotional Faces Picture Set (NIMH‐ChEFS): a new set of children's facial emotion stimuli. *International journal of methods in psychiatric research*, *20*(3), 145-156.

Erev, I., Wallsten, T. S., & Budescu, D. V. (1994). Simultaneous over- and underconfidence: The role of error in judgment processes. *Psychological Review, 101*, 519–528.

Essau, C.A., Sasagawa, S., & Frick, P.J. (2006). Callous-unemotional traits in a community sample of adolescents. *Assessment, 13,* 454-469.

Evangelista, N. M. Owens, J. S., Golden, C. M., & Pelham Jr, W. E. (2008). The Positive Illusory Bias: Do inflated self-perceptions in children with ADHD generalize to perceptions of others? *Journal of Abnormal Child Psychology*, *36,* 779-791.

Fabiano, G. A., Pelham,W. E.,Waschbusch, D., Gnagy, E. M., Lahey, B. B., Chronis, A. M., et al. (2006). A practical measure of impairment: Psychometric properties of the impairment rating scale in samples of children with attention deficit hyperactivity disorder and two school-based samples. *Journal of Clinical Child and Adolescent Psychology, 35,* 369–

385.

Fabiano, G. A., Schatz, N. K., Hulme, K. F., Morris, K. L., Vujnovic, R. K., Willoughby, M. T., ... & Pelham, W. E. (2015). Positive bias in teenage drivers with ADHD within a simulated driving task. *Journal of attention disorders*, 1087054715616186.

Faraone, S. V, Asherson, P., Banaschewski, T., Biederman, J., Ramos-quiroga, J. A., Rohde, L. A., … Franke, B. (2015). Attention-deficit/hyperactivity disorder. *Nature Reviews, 1*. http://doi.org/10.1038/nrdp.2015.20

Fefer, S. A., Ogg, J. A., & Dedrick, R. F. (2015). Use of polynomial regression to investigate biased self-perceptions and ADHD symptoms in young adolescents. *Journal of Attention Disorders*, 1087054715573993.

Fiehler, K., Ullsperger, M., & von Cramon, D. Y. (2004). Neural correlates of error detection and error correction: Is there a common neuroanatomical substrate? *European Journal of Neuroscience, 19*(11), 3081–3087.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human perception and performance*, *3*(4), 552.

Fisher, P. W., Lucas, L., Lucas, C., Sarsfield, Shaffer, D. (2006). *DISC Interviewer Manual*. Columbia University. Retrieved from http://www.cdc.gov/nchs/data/nhanes/limited_access/interviewer_manual.pdf

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: a latent-variable analysis. *Journal of Experimental Psychology: General*, *133*(1), 101.

Garavan, H., Ross, T. J., Murphy, K., Roche, R. A. P., & Stein, E. A. (2002). Dissociable executive functions in the dynamic control of behaviour: Inhibition, error detection, and correction. *NeuroImage, 17,* 1820–1829.

Gerdes, A. C., Hoza, B., & Pelham, W. E. (2003). Attention-deficit/hyperactivity disordered boys' relationships with their mothers and fathers: Child, mother, and father perceptions. *Development and Psychopathology*, *15*(02), 363-382.

Gifford-Smith, M. E., & Brownell, C. A. (2003). Childhood peer relationships: Social acceptance, friendships, and peer networks. *Journal of School Psychology, 41*, 235–284.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: a Brunswikian theory of confidence. *Psychological Review*, *98*(4), 506–28. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1961771

Gillstrom, A., & Ronnberg, J. (1995). Comprehension calibration and recall prediction accuracy of texts: Reading skill, reading strategies, and effort. *Journal of Educational Psychology, 87,* 545–558. doi:10.1037/0022-0663.87.4.545.

Greene, R. W., Biederman, J., Faraone, S. V., Monuteaux, M. C., Mick, E., & DuPre, E. P. (2001). Social impairment in girls with ADHD: Patterns, gender comparisons, and correlates. *Journal of American Academy of Child & Adolescent Psychiatry*, *40*, 704-710.

Grenell, M. M., Glass, C. R., & Katz, K. S. (1987). Hyperactive children and peer interaction: Knowledge and performance of social skills. *Journal of Abnormal Child Psychology, 15*(1), 1-13.

Herba, C. M., Landau, S., Russell, T., Ecker, C., & Phillips, M. L. (2006). The development of emotion-processing in children: Effects of age, emotion, and intensity. *Journal of Child Psychology and Psychiatry, 47*, 1098–1106. doi:10.1111/j.1469-7610.2006.01652.x

Hertenstein, M. J., & Campos, J. J. (2004). The retention effects of an adults emotional displays on infant behavior. *Child Development*, *75*, 595–613. doi:10.1111/j.1467-8624.2004.00695.x

Hervey, A. S., Epstein, J. N., & Curry, J. F. (2004). Neuropsychology of adults with attention-deficit/hyperactivity disorder: a meta-analytic review. *Neuropsychology*, *18*(3), 485.

Hinshaw, S. P., & Erhardt, D. (1992). Intervention for social competence and social skill. *Child and Adolescent Psychiatric Clinics of North America*, *1*(2), 539-552.

Hinshaw, S. P., Henker, B., & Whalen, C. K. (1984). Cognitive-behavioral and pharmacologic interventions for hyperactive boys: Comparative and combined effects. *Journal of Consulting and Clinical Psychology*, *52*(5), 739.

Hoch, S. J. (1985). Counterfactual reasoning and accuracy in predicting personal events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 719–731.

Hoza, B. (2007). Peer functioning in children with ADHD. *Journal of Pediatric Psychology, 32*, 655-663. doi:10.1093/jpepsy/jsm024

Hoza, B., Gerdes, A. C., Hinshaw, S. P., Arnold, L. E., Pelham Jr, W. E., Molina, B. S., ... & Odbert, C. (2004). Self-perceptions of competence in children with ADHD and comparison children. *Journal of consulting and clinical psychology*, *72*(3), 382.

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.

Hoza, B., Pelham, W. E., Milich, R., Pillow, D., & McBride, K. (1993). The self-perceptions and attributions of attention deficit hyperactivity disordered and nonreferred boys. *Journal of Abnormal Child Psychology, 21*, 271–286.

Hoza, B., Waschbusch, D. A., Owens, J. S., Pelham, W. E., & Kipp, H. (2001). Academic task persistence of normally achieving ADHD and control boys: Self-evaluations and attributions. *Journal of Consulting and Clinical Psychology, 69*, 271-283.

Hoza, B., Waschbusch, D. A., Pelham, W. E., Molina, B. S. G., & Milich, R. (2000). Attention-deficit/hyperactivity disordered and control boys' responses to social success and failure. *Child Development, 71,* 432-446.

Jia, M., Jiang, Y., & Mikami, A. Y. (2015). Positively biased self-perceptions in children with ADHD: Unique predictor of future maladjustment. *Journal of Abnormal Child Psychology*, 1-12.

Joseph, L. M., & Eveleigh, E. L. (2011). A review of the effects of self-monitoring on reading performance of students with disabilities. *The Journal of Special Education, 45,* 43-53. doi:10.1177/ 022466909349145

Juslin, P., Winman, A., & Olsson, H. (2000). Naïve empiricism and dog- matisim in confidence research: a critical examination of the hard-easy effect. *Psychology Review, 107*, 384–396

Kaufman, A. S., & Kaufman, N. L. (2004b). *Kaufman Brief Intelligence Test, Second Edition*. Bloomington, MN: Pearson, Inc.

Keren, G. (1991). Calibration and probability judgments: Conceptual and methodological issues. *Acta Psychologica*, *77*(3), 217-273.

Kimball, D. R., & Metcalfe, J. (2003). Delaying judgments of learning affects memory, not metamemory. *Memory & Cognition*, *31*(6), 918-929.

Klayman, J. (1995).Varieties of confirmation bias. In J. Busemeyer,R. Hastie, & D.L. Medin

(Eds.), *Psychology of learning and motivation: Vol. 32. Decision making from a cognitive perspective* (pp. 365–418). New York: Academic Press.

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review*, *96*, 674-689.

Kuhn, D. (2014). Developmental Model of Critical Thinking. *Educational Researcher*, *28*(2), 16–25.

Klayman, J., & Soll, J. (1999). Overconfidence: It Depends on How, What, and Whom You Ask. *Organizational Behavior and Human Decision Processes*, *79*(3), 216–247.

Knouse, L. E., Paradise, M. J., & Dunlosky, J. (2006). Does ADHD in adults affect the relative accuracy of metamemory judgments? *Journal of Attention Disorders*, *10*(2), 160–170.

Koriat, A., & Ackerman, R. (2010). Choice latency as a cue for children’s subjective confidence in the correctness of their answers. *Developmental Science*, *13*(3), 441–53.

Kreutzer, M., Leonard, C., & Flavell, J. (1975). An interview study of children's knowledge about memory. *Society for Research in Child Development Monographs, 40* (Serial No. 159).

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one’s own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*, 1121–1134. doi:10.1037/0022-3514.77.6.1121

Krueger, C. E., Rosen, H. J., Taylor, H. G., Espy, K. A., Schatz, J., Rey-Casserly, C., & Kramer, J. H. (2011). Know thyself: real-world behavioral correlates of self-appraisal accuracy.

The Clinical Neuropsychologist, 25(5), 741-756.

Larrick, R. P. (2004). Debiasing. In D. K. Koehler, & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (pp. 316–337). Oxford, England: Blackwell.

Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin*, *125*(2), 255–275.

Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? *Organizational Behaviour and Human Decision Processes.*, *20*(2), 159–183.

Lichtenstein, S., & Fischhoff, B. (1980). Training for calibration. *Organizational Behavior and Human …*, *171*, 149–171. Retrieved from http://www.sciencedirect.com/science/article/pii/0030507380900525

Lin, L. M., Zabrucky, K. M., & Moore, D. (2001). Effects of text difficulty and adults' age on relative calibration of comprehension. *The American Journal of Psychology*, *115*(2), 187-198.

Loge, D. V., Staton, R. D., & Beatty, W. W. (1990). Performance of children with ADHD on tests sensitive to frontal lobe dysfunction. *Journal of the American Academy of Child & Adolescent Psychiatry*, *29*(4), 540-545.

Luby, J. L., Belden, A., Sullivan, J., & Spitznagel, E. (2007). Preschoolers’ contribution to their diagnosis of depression and anxiety: Uses and limitations of young child self-report of symptoms. *Child Psychiatry and Human Development*, *38*(4), 321-338.

Lundeberg, M., & Mohan, L. (2010). Context matters: Gender and cross-cultural differences in confidence. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of Metacognition in Education* (pp. 221–239). London and New York: Routledge.

Maki, R. H., Shields, M., Wheeler, A. E., & Zacchilli, T. L. (2005). Individual differences in absolute and relative metacomprehension accuracy. *Journal of Educational Psychology*, *97*, 723–731. doi:10.1037/ 0022-0663.97.4.723.

Mann, L., Harmoni, R., & Power, C. (1989). Adolescent decision-making: the development of competence. *Journal of Adolescence*, *12*(3), 265–78. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2687339

Mäntylä, T., Still, J., Gullberg, S., & Del Missier, F. (2012). Decision making in adults with ADHD. *Journal of Attention Disorders*, *16*(2), 164–73.

Marton, I., Wiener, J., Rogers, M., & Moore, C. (2015). Friendship characteristics of children with ADHD. *Journal of attention disorders, 19*(10), 872-881.

Matthys, W., Cuperus, J. M., & Engeland, H. Van. (1999). Deficient Social Problem-Solving in Boys With ODD/CD, With ADHD, and With Both Disorders. *Journal of the American Academy of Child & Adolescent Psychiatry*, *38*(3), 311–321.

Mauboussin, M. J. & Callahan, D. (2015). IQ versus RQ: Differentiating Smarts from Decision-Making Skills. *Global Financial Strategies.*

McAlpine, C., Singh, N. N., Ellis, C. R., Kendall, K. A., & Hampton, C. (1992). Enhancing the ability of adults with mental retardation to recognize facial expressions of emotion. *Behavior Modification*, *16*(4), 559-573.

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.

Milich, R., & Dodge, K.A. (1984). Social information processing in child psychiatry populations. *Journal of Abnormal Child Psychology, 12,* 471–489

Montirosso, R., Peverelli, M., Frigerio, E., Crespi, M., & Borgatti, R. (2010). The development of dynamic facial expression recognition at different intensities in 4- to 18-year-olds. *Social Development*, *19*(1), 71–92. doi:10.1111/j.1467-9507.2008.00527.x

National Health and Medical Research Council (NHMRC, 2009). Attention Deficit Hyperactivity Disorder: The NICE guidelines on diagnosis and management of ADHD in children, young people and adults (Vol. National Clinical Practice Guidelines). London: The British Psychological Society and The Royal College of Psychiatrists. Retrieved from .http://www.nice.org.uk/nicemedia/pdf/ADHDFullGuideline.pdf.

Nelson, T. O. (1996). Gamma is a Measure of the Accuracy of Predicting Performance on One Item Relative to Another Item, not of the Absolute Performance on an Individual Item Comments on Schraw (1995). *Applied Cognitive Psychology*, *10*(3), 257-260.

Nelson, T. O., Narens, L., & Dunlosky, J. (2004). A revised methodology for research in metamemory: Pre- judgment recall and monitoring (PRAM). *Psychological Methods, 9*,53–69. doi:10.1037/1082- 989X.9.1.53.

Newman, R. S. (1984). Children’s numerical skill and judgments of confidence in estimation. *Journal of Experimental Child Psychology*, *37*(1), 107–123.

Nijmeijer, J.S., Minderaa, R.B., Buitelaar, J.K., Mulligan, A., Hartman, C.A., Hoekstra, P.J., 2008. Attention-deficit/hyperactivity disorder and social dysfunctioning. *Clin. Psychol.*

Rev. 28 (4), 692–708.

O'Connell, R. G., Bellgrove, M. A., Dockree, P. M., Lau, A., Hester, R., Garavan, H., ... & Robertson, I. H. (2009). The neural correlates of deficient error awareness in attention-deficit hyperactivity disorder (ADHD).*Neuropsychologia*, *47*(4), 1149-1159.

Olson, D., & Astington, J. (1993). Thinking about thinking: Learning how to take statements and hold beliefs. *Educational Psychologist*, *28*(1), 7-23.

Osterlind, S. J. (2006). *Modern measurement: Theory, practice, and applications of mental appraisal*. Upper Saddle River: Prentice Hall.

Owens, J. S., & Hoza, B. (2003). The role of inattention and hyperactivity/impulsivity in the positive illusory bias. *Journal of Consulting and Clinical Psychology*, *71*(4), 680.

Parker, A. M., & Fischhoff, B. (2005). Decision-making compe- tence: External validation through an individual-differences approach. *Journal of Behavioral Decision Making, 18,* 1–27.

Philippot, P., & Feldman, R.S. (1990). Age and social competence in preschoolers' decoding of facial expression. *British Journal of Social Psychology, 29* (1), 43–54.

Pliszka S. (2007). Practice parameter for the assessment and treatment of children and adolescents with attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 46,* 894–921.

Polanczyk, G., de Lima, M. S., Horta, B. L., Biederman, J., & Rohde, L. A. (2007). The worldwide prevalence of ADHD: A systematic review and metaregression analysis. *American Journal of Psychiatry, 164*(6), 942–948. http://dx.doi.org/10.1176/appi.ajp.164.6.942

Reitan, R. M. (1958). Validity of the trail making test as an indicator of organic brain damage. *Perceptual Motor Skills*, *8*, 271-276.

Schachar, R. J., Chen, S., Logan, G. D., Ornstein, T. J., Crosbie, J., Ickowicz, A., et al. (2004). Evidence for an error monitoring deficit in Attention Deficit Hyperactivity Disorder. *Journal of Abnormal Child Psychology, 32*(3), 285–293.

Schäfer, M. (2011). Delay of gratification and the ability to self-regulation in preschool children. Unpublished master's thesis, Goethe University Frankfurt, Frankfurt, Germany

Scheithauer, M. C., & Kelley, M. L. (2014). Self-monitoring by college students with ADHD: The impact on academic performance. *Journal of Attention Disorders*. http://doi.org/10.1177/1087054714553050

Schneider, W., Visé, M., Lockl, K., & Nelson, T. O. (2000). Developmental trends in children’s memory monitoring: Evidence from a judgment-of-learning task. *Cognitive Development*, *15*, 115–134. Retrieved from http://www.sciencedirect.com/science/article/pii/S0885201400000241

Scholtens, S., Diamantopoulou, S., Tillman, C. M., & Rydell, A. (2012). Effects of symptoms of ADHD, ODD, and Cognitive Functioning on Social Acceptance and the Positive Illusory Bias in Children. *Journal of Attention Disorders*, *16*(8), 685-696.

Schraw, G. (2008). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and Learning*, *4*(1), 33–45. http://doi.org/10.1007/s11409-008-9031-3

Schraw, G. (2010). Measuring metacognitive judgments. In D. J. Hacker, J. Dunlosky, & A. C.

Graesser (Eds.), *Hand* (pp. 415–429). London and New York.

Schulz, K. P., Bédard, A.-C. V, Fan, J., Clerkin, S. M., Dima, D., Newcorn, J. H., & Halperin, J. M. (2014). Emotional bias of cognitive control in adults with childhood attention-deficit/hyperactivity disorder. *NeuroImage: Clinical*, *5*, 1–9.

Seidman, L. J. (2006). Neuropsychological functioning in people with ADHD across the lifespan. *Clinical Psychology Review, 4*, 466–485.

Serrano, V. J., Owens, J. S., & Hallowell, B. (2015). Where children with ADHD direct visual attention during emotion knowledge tasks relationships to accuracy, response rime, and ADHD symptoms. *Journal of Attention Disorders*, doi: 1087054715593632.

Shaffer, D., Fisher, P., Lucas, C. P., Dulcan, M. K., & Schwab-Stone, M. E. (2000). NIMH Diagnostic Interview Schedule for Children Version IV (NIMH DISC-IV). Description , Differences From Previous Versions , and Reliability of Some Common Diagnoses. *Journal of the American Academy of Child & Adolescent Psychiatry*, *39*(1), 28–38.

Shapiro, E. G., Hughes, S. J., August, G. J., & Bloomquist, M. L. (1993). Processing of emotional information in children with attention-deficit hyperactivity disorder. *Developmental Neuropsychology, 9*(3–4), 207–224.

Simon, E. W., Rosen, M., Grossman, E., & Pratowski, E. (1995). The relationships among facial emotion recognition, social skills, and quality of life. *Research in Developmental Disabilities, 16*(5), 383–391.

Singh, S. D., Ellis, C. R., Winton, A. S., Singh, N. N., Leung, J. P., & Oswald, D. P. (1998).

Recognition of facial expressions of emotion by children with attention-deficit hyperactivity disorder. *Behav Modif*, *22*(2), 128–142.

Smith, B. H., Barkley R. A., Shapiro, C. J. (2007). Attention-Deficit/Hyperactivity Disorder. In E. J. Mash & R. A. Barkley (Eds), *Assessment of Childhood Disorders* (4th ed.) (53-131). Guilford Press; New York.

Souchay, C., Isingrini, M., Clarys, D., Taconnat, L., & Eustache, F. (2004). Executive functioning and judgment-of-learning versus feeling-of-knowing in older adults. *Experimental Aging Research*, *30*(1), 47-62.

Stanovich, K. E., & West, R. F. (2000). Individual difference in reasoning: Implication for the rationality debate. *Behavioural and Brain Sciences, 23*, 645–665. doi:10.1017/S0140525X00003435

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd ed.). New York: Oxford University Press.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18,* 643-662.

Tannock, R. (1998). Attention deficit hyperactivity disorder: Advances in cognitive, neurobiological, and genetic research. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 39*(1), 65–99.

Taylor, E., Dopfner, M., Sergeant, J., Asherson, P., Banaschewski, T., Buitelaar, J.,…Zuddas, A. (2004) European clinical guidelines for hyperkinetic disorder – first upgrade. European Child and Adolescent Psychiatry 13(Suppl 1):7-30.

Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology*, *95*(1), 66–73. http://doi.org/10.1037/0022-0663.95.1.66

Thiede, K. W., Dunlosky, J., Griffin, D., & Wiley, J. (2005). Understanding the delayed-keyword effect on metacomprehension accuracy. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *31*, 1267–1280. doi:10.1037/0278-7393.31.6.1267.

Thompson, V. A. , Prowse Turner, J. A. , & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive Psychology*, *63*(3), 107–40. doi:10.1016/j.cogpsych.2011.06.001

Thompson, W. B. (1999). Individual differences in memory-monitoring accuracy. *Learning and Individual Differences*, *11*(4), 365-376.

Toplak, M. E., Connors, L., Shuster, J., Knezevic, B., & Parks, S. (2008). Review of cognitive, cognitive-behavioral, and neural-based interventions for Attention-Deficit/Hyperactivity Disorder (ADHD). *Clinical psychology review*,*28*(5), 801-823.

Uekermann, J., Kraemer, M., Abdel-Hamid, M., Schimmelmann, B. G., Hebebrand, J., Daum, I., Wiltfang, J., & Kis, B. (2010). Social cognition in attention-deficit hyperactivity disorder (ADHD). *Neuroscience & Biobehavioral Reviews*, *34*(5), 734–743. http://doi.org/10.1016/j.neubiorev.2009.10.009

Van der Linden, D., Frese, M., & Meijman, T. F. (2003). Mental fatigue and the control of cognitive processes: effects on perseveration and planning. *Acta Psychologica*, *113*(1), 45-65.

van Meel, C. S., Heslenfeld, D. J., Oosterlaan, J., & Sergeant, J. A. (2007). Adaptive control deficits in attention-deficit/hyperactivity disorder (ADHD): the role of error processing. *Psychiatry Research*, *151*(3), 211-220.

Veenman, M. V. J., Van Hout-Walters, B. H. A., & Afflerbach, P. (2006). Metacognition and learning: conceptual and methodological considerations. *Metacognition and Learning, 1,* 3–14. doi:10.1007/ s11409-006-6893-0.

Weaver, C. A. (1990). Constraining factors in the calibration of comprehension. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 16,* 214–222. doi:10.1037/0278-7393.16.2.214.

Webber, J., Scheuermann, B., McCall, C., & Coleman, M. (1993). Research on self-monitoring as a behavior management technique in special education classrooms a descriptive review. *Remedial and Special Education*, *14*(2), 38-56.

Weller, J. A., Levin, I. P., Rose, J. P., & Bossard, E. (2012). Assessment of decision-making competence in preadolescence. *Journal of Behavioral Decision Making*, *25*(4), 414-426.

Whalen, C. K., & Henker, B. (1985). The social worlds of hyperactive (ADHD) children. *Clinical Psychology Review, 5,* 447–478.

Willcutt, E. G., Nigg, J. T., Pennington, B. F., Solanto, M. V., Rohde, L. A., Tannock, R., et al. (2012). Validity of DSM-IV attention-deficit/hyperactivity disorder dimensions and subtypes. *Journal of Abnormal Psychology, 121*(4), 991.

Wolfe, R. N., & Grosch, J. W. (1990). Personality correlates of confidence in one's decisions. *Journal of Personality*, *58*(3), 515–534. http://doi.org/10.1111/j.1467-6494.1990.tb00241.x

Yates, J. F. (1990). *Judgment and decision making*. Englewood Cliffs, NJ: Prentice Hall.

Appendix A: Battery Versions

Child Checklist -VA

________*Assent Form*

________ *Kaufman Brief Intelligence Test, Second Edition*

________ *Kaufman Brief Intelligence Test, Second Edition_Cognitive Effort*

________*The Trailmaking Test*

________*The Trailmaking Test_Cognitive Effort*

________*The Stroop Task*

________*The Stroop Task_Cognitive Effort*

________*ADHD Performance Task*

________*ADHD Performance Task_Cognitive Effort*

________*General Knowledge Confidence Task*

________*General Knowledge Confidence Task_Cognitive Effort*

________*Child Social Situation Confidence Task*

________*Child Social Situation Confidence Task_Cognitive Effort*

________*Emotion Recognition Confidence Task*

________*Emotion Recognition Confidence Task_Cognitive Effort*

Notes:

Child Checklist - VB

________*Assent Form*

________*General Knowledge Confidence Task*

________*General Knowledge Confidence Task_Cognitive Effort*

________*Child Social Situation Confidence Task*

________*Child Social Situation Confidence Task_Cognitive Effort*

________*Emotion Recognition Confidence Task*

________*Emotion Recognition Confidence Task_Cognitive Effort*

________ *Kaufman Brief Intelligence Test, Second Edition*

________ *Kaufman Brief Intelligence Test, Second Edition_Cognitive Effort*

________*The Trailmaking Test*

________*The Trailmaking Test_Cognitive Effort*

________*The Stroop Task*

________*The Stroop Task_Cognitive Effort*

________*ADHD Performance Task*

________*ADHD Performance Task_Cognitive Effort*

Notes:

Appendix B: General Knowledge Task and Percent Correct by Item and Group

General Knowledge Task

Read each question aloud. Ensure child makes ratings after each question.

Say:

"Now, we are going to answer some questions. I am going to read you a question, and give you three possible answers. Please choose the answer you think is most correct. Some questions are going to be easy, and some questions are going to be hard. Just give each question your best try!"

Now, and after each item say:

"After each question I would like you to rate how confident or sure you are that you got the question right. You can make this rating anywhere from 0 – I'm not sure if I am right to 5-Good Chance I'm right to 10-I'm definitely right! (Point to line after sample item) **You can pick any number on this line, even the ones in between the words. And remember, it is OK to get a question wrong!"**

Once the child understands the confidence rating, you do not have to keep repeating this instruction. Simply prompt a confidence rating by asking, "**How confident/sure are you that you got the question right?"**

"Let's try one for practice!"

After practice item say, **"Great!"**

Record response on **General Knowledge Task Scoring** sheet.

General Knowledge Items

Item	Entire Sample % Correct	Clinical ADHD Group % Correct	TD Group % Correct
1) The opposite of dark a. Happy b. Light* c. Song	95.1	94.9	95.2
2) A type of oven than cooks food very quickly a. Microwave* b. Fridge c. Stove	72.8	74.4	71.4
3) A short coat a. Waist b. Leather c. Jacket*	60.5	64.1	57.1
4) Something you stick on an envelope a. Sticker b. Stamp* c. Flower	86.4	87.2	85.7
5) The seventh month of the year a. January b. July* c. June	76.5	66.7	85.7
6) The colour of grass a. Green* b. Orange c. Grey	100	100	100
7) Bones which protect your heart and lungs a. Skull b. Tibia c. Ribs*	82.7	79.5	85.7
8) Something you hit with a hammer a. Nail* b. Hat c. Screw	92.6	92.3	92.9
9) Somebody who works with wood a. Painter b. Carpenter* c. Electrician	87.7	89.7	85.7

10) A common garden insect Easy a. Ant* b. Uncle c. Cat	97.5	94.9	100
11) A reference book containing list of synonyms a. Thesaurus* b. Menu c. Dictionary	45.7	41	50
12) A noise that warns you of danger a. Car b. Alarm* c. Computer	95.1	92.3	97.6
13) The opposite of smooth a. Rough* b. Long c. Cone	96.3	92.3	100
14) A picture that shows your bones a. Painting b. Skeleton c. Xray*	97.5	100	95.2
15) A colourless, tasteless, odourless gas a. Water b. Oxygen* c. Sodium	72.8	64.1	81.0
16) A red insect with black spots a. Fly b. Ladybug* c. Lobster	97.5	94.9	100
17) A sea creature with 8 legs a. Snail b. Fish c. Octopus*	100	100	100
18) Extremely hungry a. Thirsty b. Famished* c. Growling	48.1	51.3	45.2
19) Cars, buses, trucks, vans, etc., a. Animals b. Foods	100	100	100

c. Vehicles*			
20) Something that attracts metal a. Steel b. Tree c. Magnet*	98.8	97.4	100
21) To shout a. Yell* b. Whisper c. Yelp	92.6	94.9	90.5
22) A wild horse from Africa a. Elephant b. Giraffe c. Zebra*	87.7	87.2	88.1
23) The joint that connects a foot with a leg a. Elbow b. Ankle* c. Shoulder	95.1	94.9	95.2
24) A martial art a. Karaoke b. Kick c. Karate*	87.7	89.7	85.7
25) A coloured handkerchief worn around the head or neck a. Bandana* b. Hat c. Necklace	86.4	89.7	83.3
26) The closest planet to the sun a. Mercury* b. Mars c. Venus	44.4	41	47.6
27) A large piece of rock a. Pebble b. Stone c. Bedrock*	54.3	61.5	47.6
28) A fast breed of dog a. Greyhound* b. Rottweiler c. Pug	50.6	56.4	45.2
29) To join together a. Split b. Unite* c. Joint	70.4	71.8	69

30) A floating mass of ice a. Ship b. Plate c. Iceberg*	98.8	100	97.6
31) The science of matter and energy a. Mathematics b. Biology c. Physics*	55.6	61.5	50
32) A brother of your father or mother a. Aunt b. Uncle* c. Nephew	82.7	79.5	85.7
33) Something you do when you feel tired a. Yawn* b. Play c. Smile	97.5	94.9	100
34) A mountain through which lava erupts a. Hill b. Volcano* c. Cave	98.8	97.4	100
35) A citrus fruit a. Orange* b. Kiwi c. Banana	49.4	43.6	54.8
36) Microsoft's gaming console a. Xbox* b. Playstation c. Wii	61.7	51.3	71.4

Appendix C: Emotion Recognition Task and Percent Correct by Item and Group

Emotion Recognition Task Script

"Now, we are going to do a faces task! You will see a photo of a child at the top of the page and a list of five feelings at the bottom of the page. You need to pick the name of the feeling that best describes the feelings that you see. Sometimes, it may be hard to pick just one feeling – when that happens, just choose the feeling you think best describes the child's face."

"Let's try one for practice! Look at this face (point to the top picture). **Pick a feeling at the bottom of the page that matches the feeling this child has on his/her face."** (Repeat for each item).

After child responds, point to the confidence rating scale and say, **"Rate how confident or sure you are that you are right anywhere from 0 – I'm not sure know if I am right to 5-Good Chance I'm right to 10-I'm definitely right! You can pick any number on this line. You can pick any number on this line, even the ones in between the words."** (Repeat after each item, do not have to keep repeating this instruction if the child understands how to make the rating).

After practice item say, **"Great! Remember, pick a feeling at the bottom of the page that matches the emotion the child has on his/her face. If you find it hard to pick the right feeling, just choose the feeling you think best describes the child's face. Some faces may be easy, and some may be hard. Just give each question your best try!"**

If child asks what feeling the child is expressing, say, **" I can't tell you what they are feeling, I want you to figure it out."** Or, **"Just take your best guess."**

Mark child's responses on the Emotion Recognition Task Scoring sheet.

Emotion Recognition Task Stimuli

Sample Item:

A) Sad

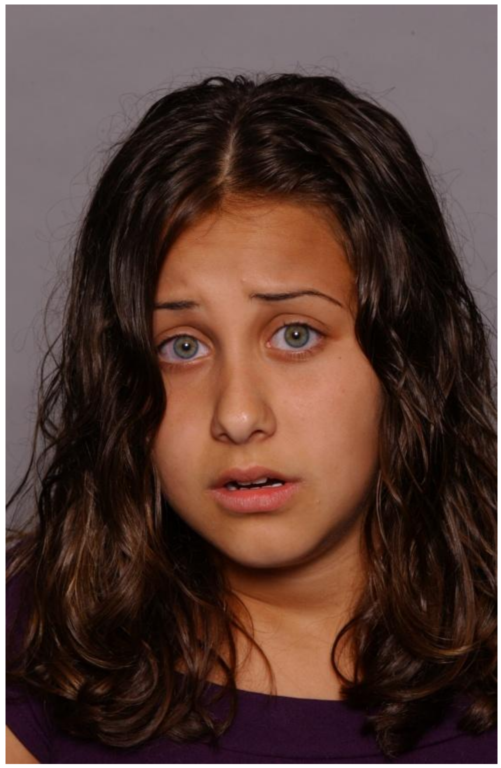
B) Happy*

C) Angry

D) Afraid

E) Neutral

Item 1:



A) Sad

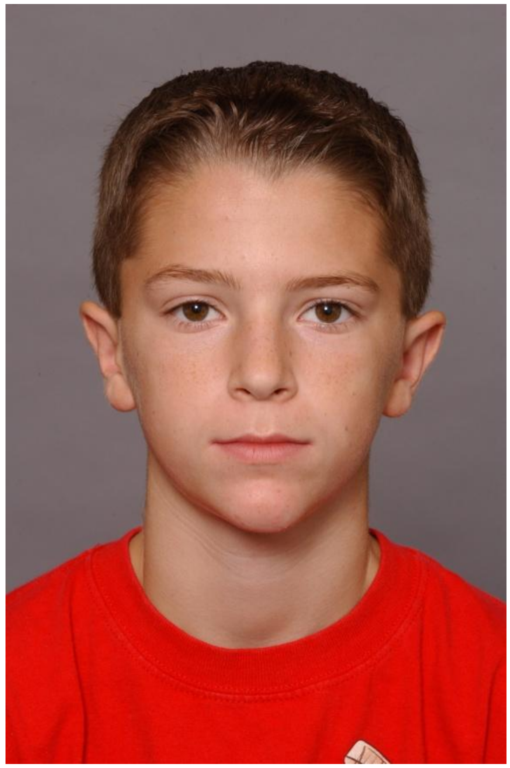
B) Happy

C) Angry

D) Afraid*

E) Neutral

Item 2:



A) Sad

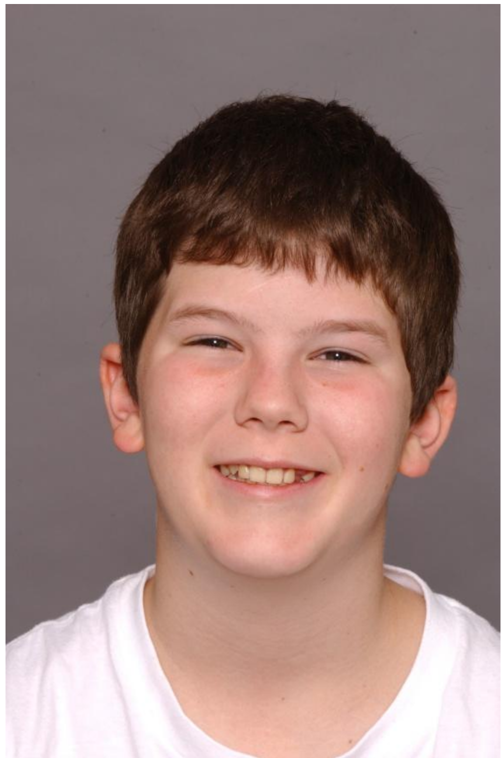
B) Happy

C) Angry

D) Afraid

E) Neutral*

Item 3:



A) Sad

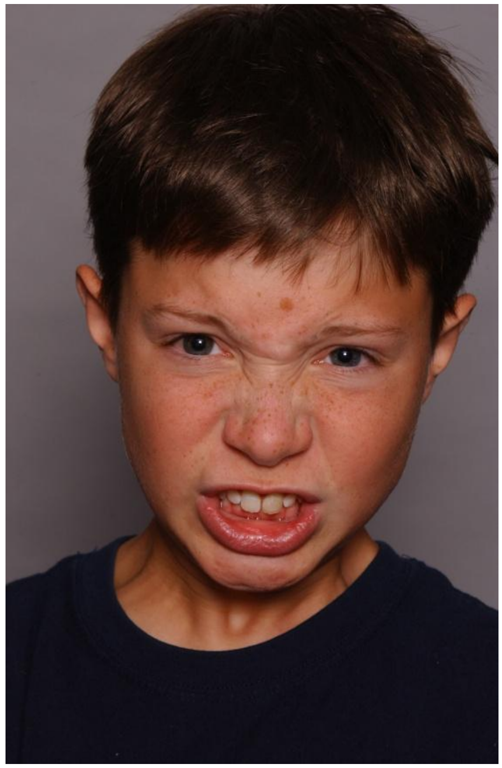
B) Happy*

C) Angry

D) Afraid

E) Neutral

Item 4:



A) Sad

B) Happy

C) Angry*

D) Afraid

E) Neutral

Item 5:

A) Sad

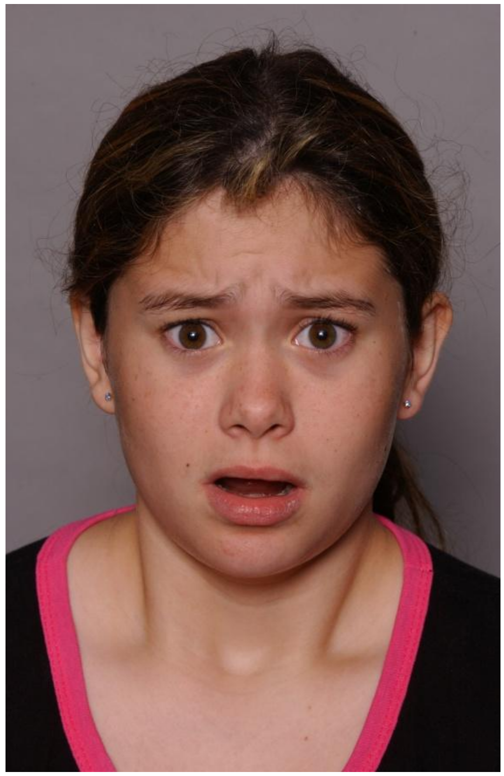
B) Happy*

C) Angry

D) Afraid

E) Neutral

Item 6:



A) Sad

B) Happy

C) Angry

D) Afraid*

E) Neutral

Item 7:

A) Sad

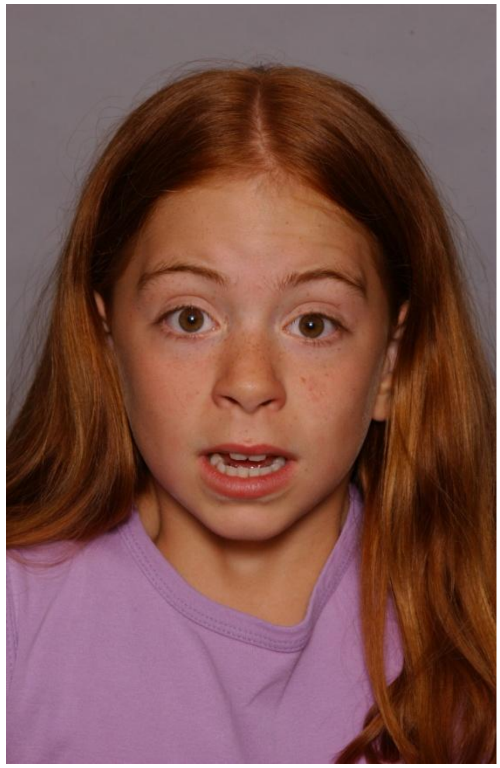
B) Happy

C) Angry

D) Afraid

E) Neutral*

Item 8:



A) Sad

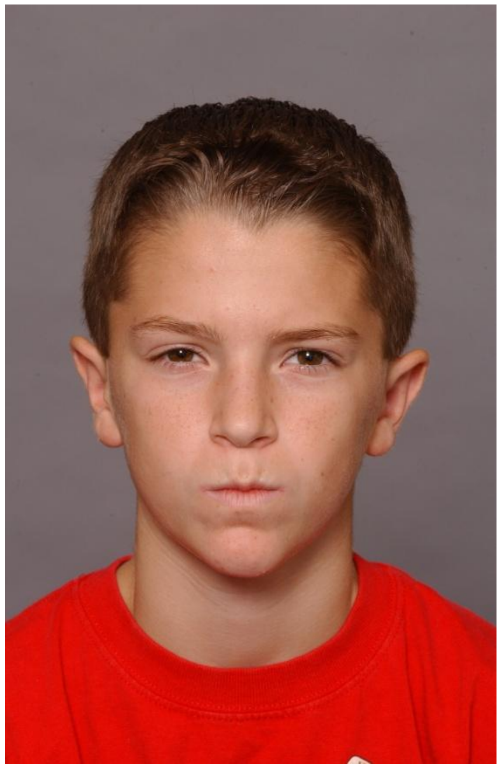
B) Happy

C) Angry

D) Afraid*

E) Neutral

Item 9:



A) Sad

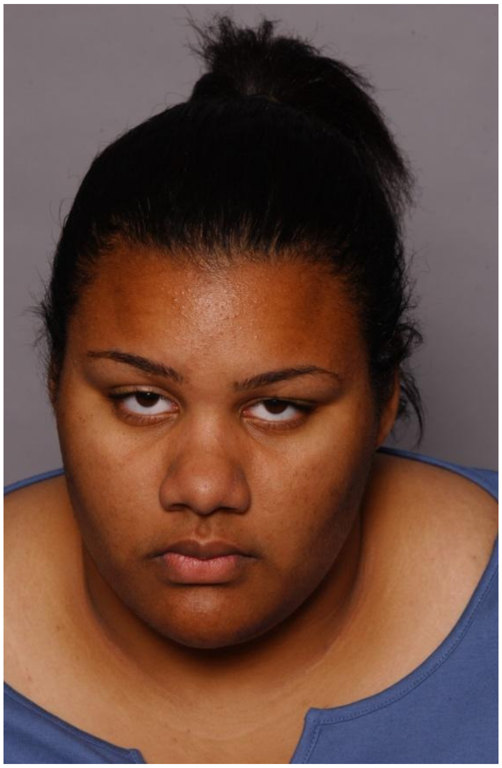
B) Happy

C) Angry*

D) Afraid

E) Neutral

Item 10:



A) Sad*

B) Happy

C) Angry

D) Afraid

E) Neutral

Item 11:

A) Sad

B) Happy

C) Angry

D) Afraid

E) Neutral*

Item 12:

A) Sad

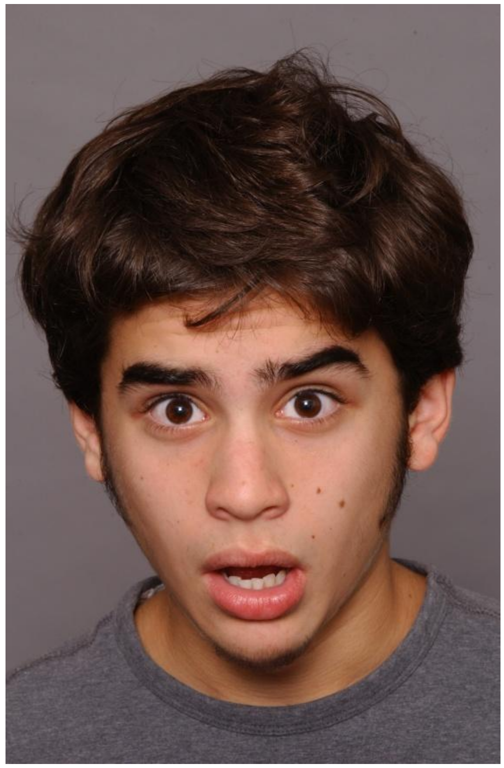
B) Happy*

C) Angry

D) Afraid

E) Neutral

Item 13:



A) Sad

B) Happy

C) Angry

D) Afraid*

E) Neutral

Item 14:

A) Sad

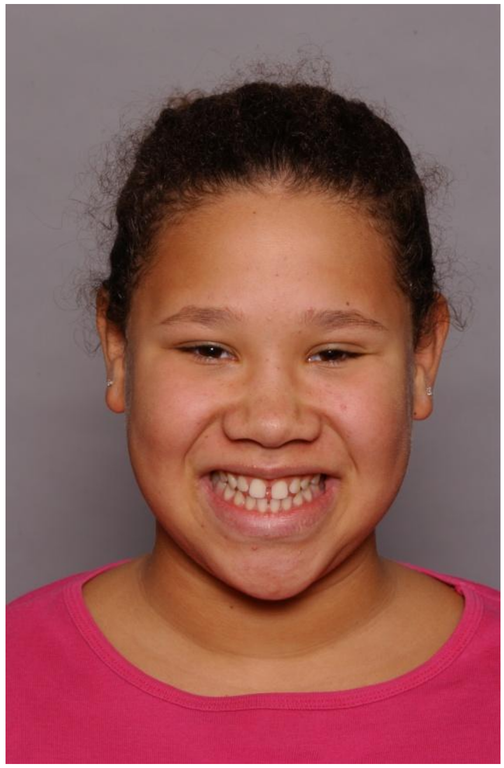
B) Happy

C) Angry

D) Afraid

E) Neutral*

Item 15:



A) Sad

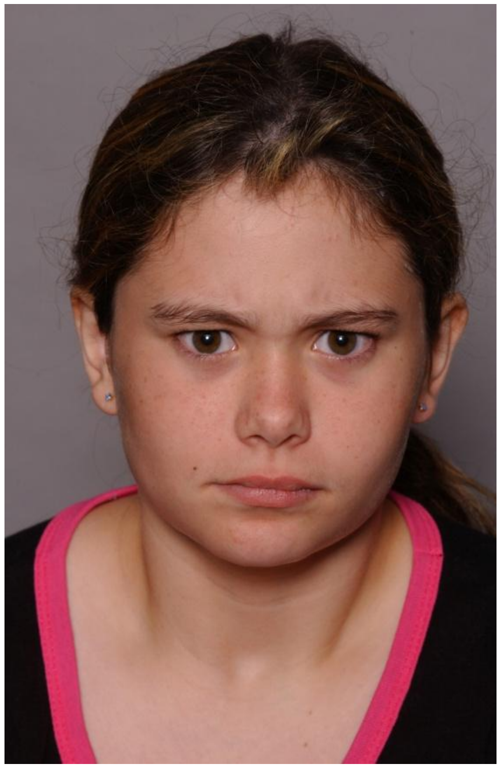
B) Happy*

C) Angry

D) Afraid

E) Neutral

Item 16:



A) Sad

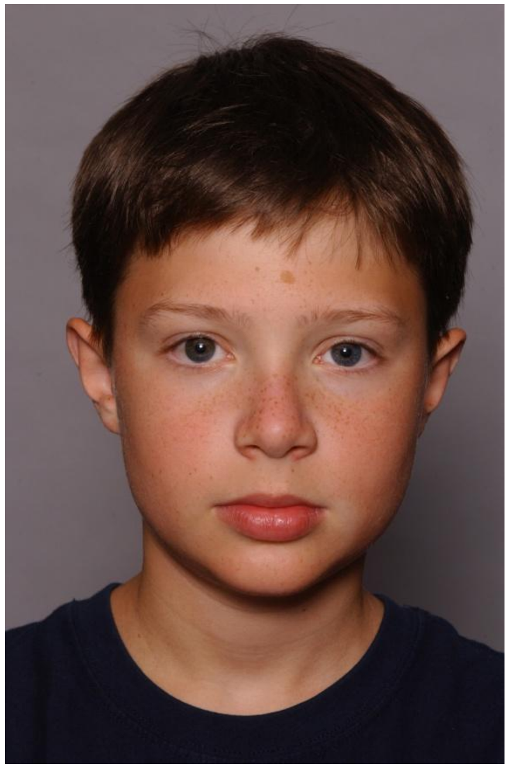
B) Happy

C) Angry*

D) Afraid

E) Neutral

Item 17:



A) Sad

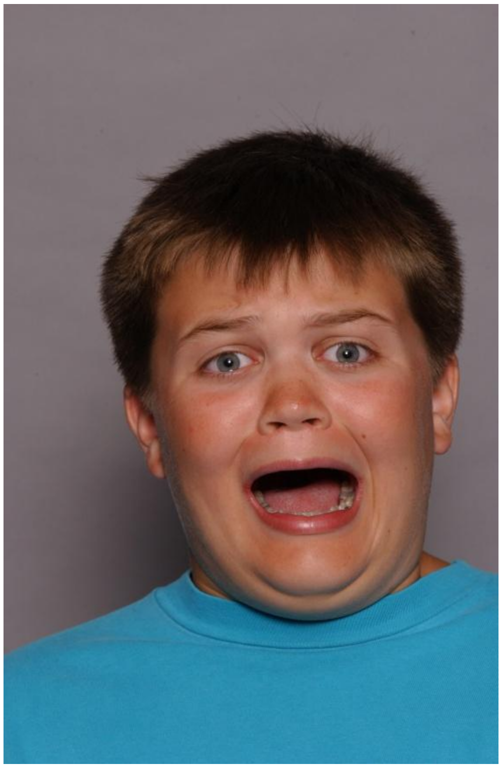
B) Happy

C) Angry

D) Afraid

E) Neutral*

Item 18:



A) Sad

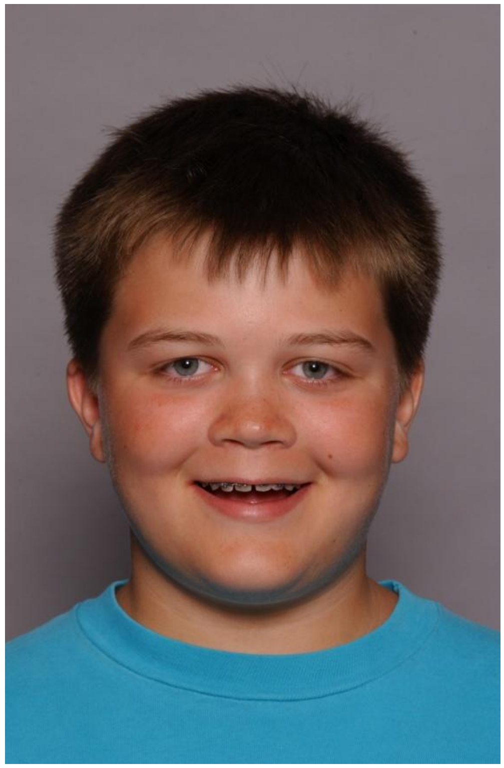
B) Happy

C) Angry

D) Afraid*

E) Neutral

Item 19:



A) Sad

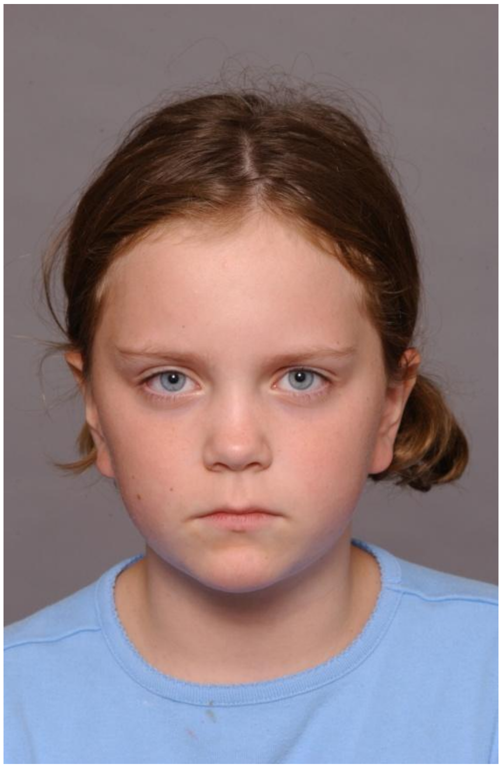
B) Happy*

C) Angry

D) Afraid

E) Neutral

Item 20:



A) Sad

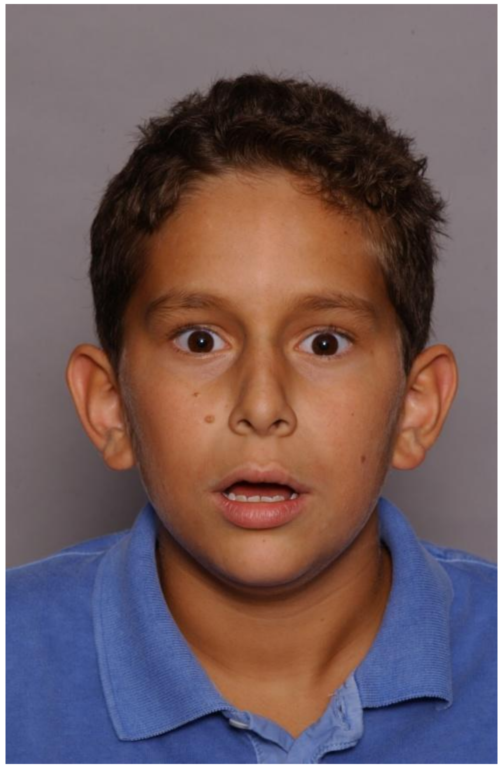
B) Happy

C) Angry

D) Afraid

E) Neutral*

Item 21:



A) Sad

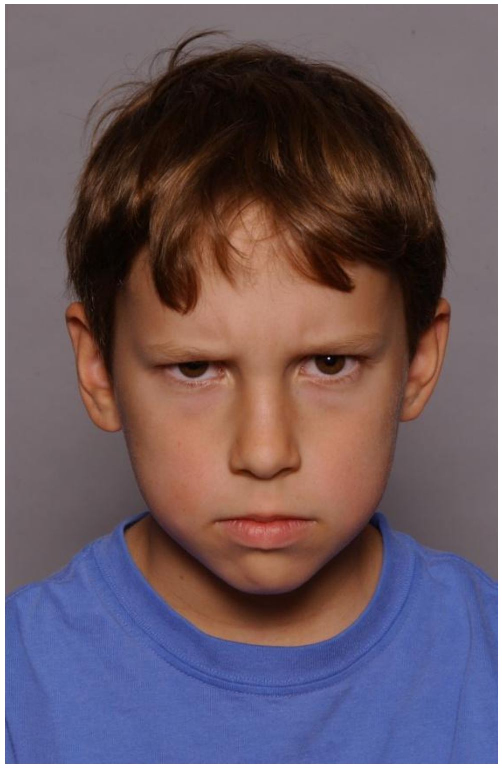
B) Happy

C) Angry

D) Afraid*

E) Neutral

Item 22:



A) Sad

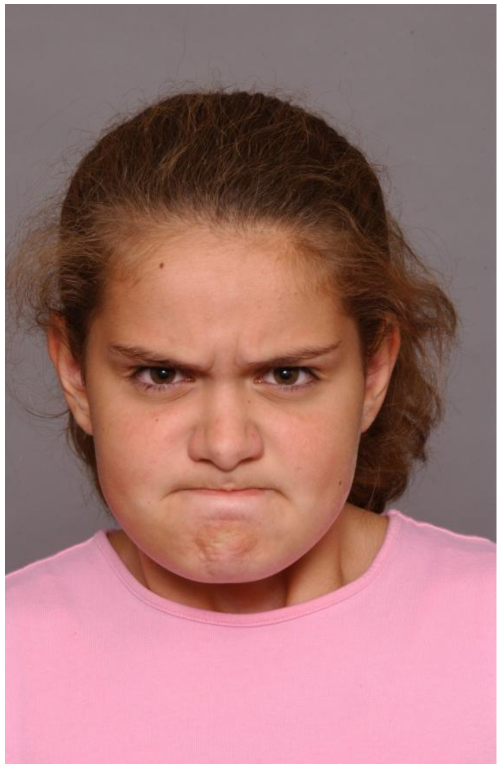
B) Happy

C) Angry*

D) Afraid

E) Neutral

Item 23:



A) Sad

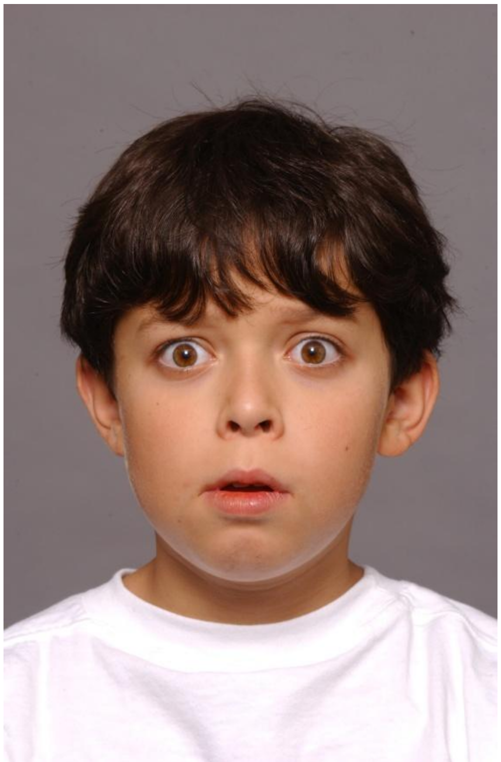
B) Happy

C) Angry*

D) Afraid

E) Neutral

Item 24:



A) Sad

B) Happy

C) Angry

D) Afraid*

E) Neutral

Item 25:

A) Sad

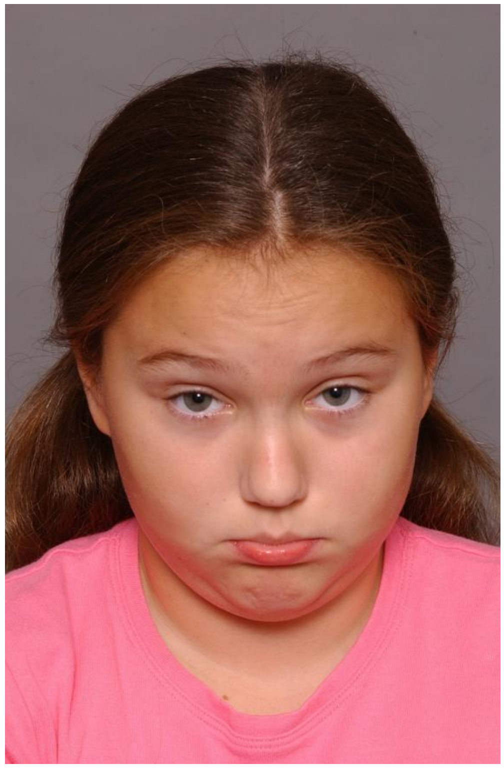
B) Happy*

C) Angry

D) Afraid

E) Neutral

Item 26:



A) Sad*

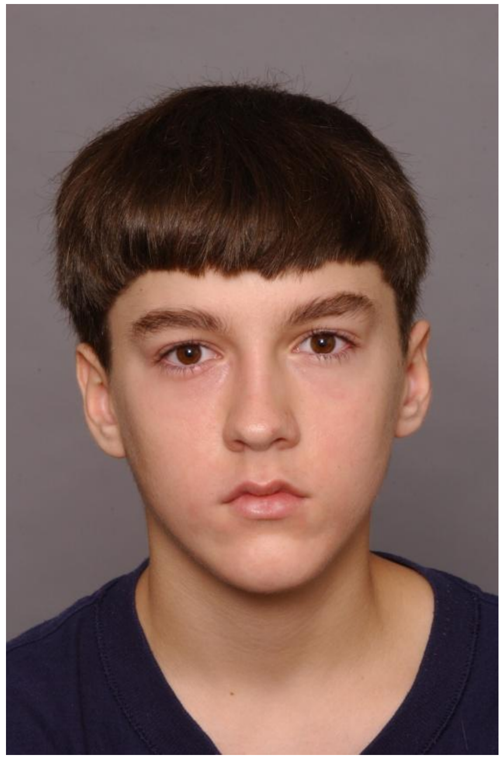
B) Happy

C) Angry

D) Afraid

E) Neutral

Item 27:



A) Sad

B) Happy

C) Angry

D) Afraid

E) Neutral*

Item 28:

A) Sad*

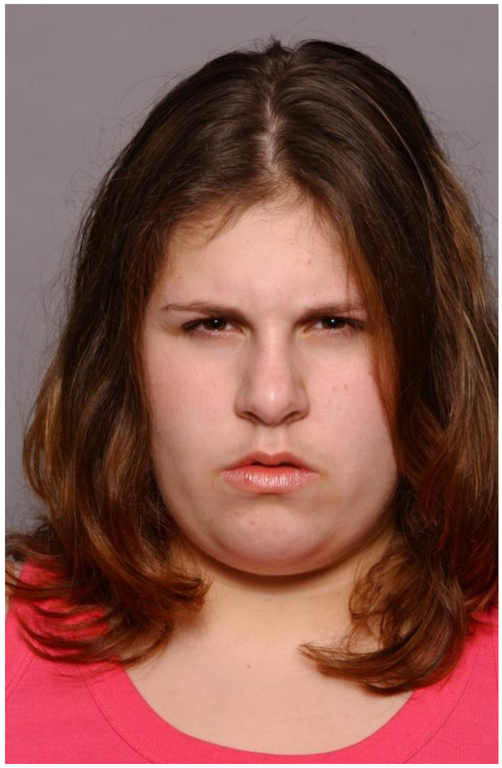
B) Happy

C) Angry

D) Afraid

E) Neutral

Item 29:



A) Sad

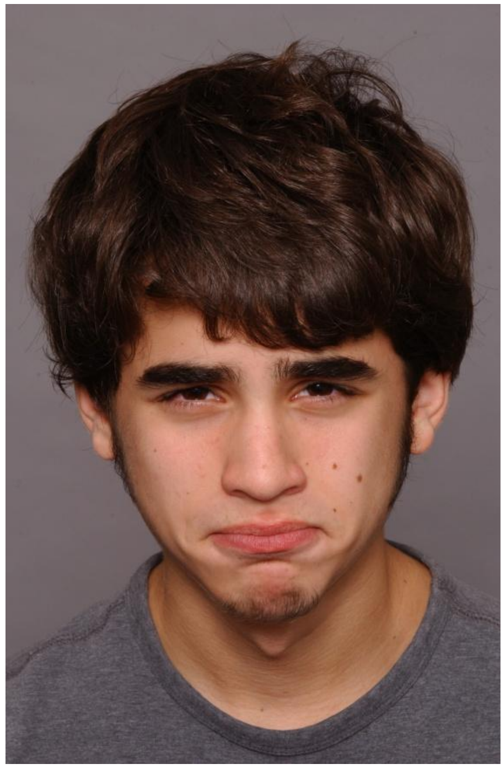
B) Happy

C) Angry*

D) Afraid

E) Neutral

Item 30:



A) Sad*

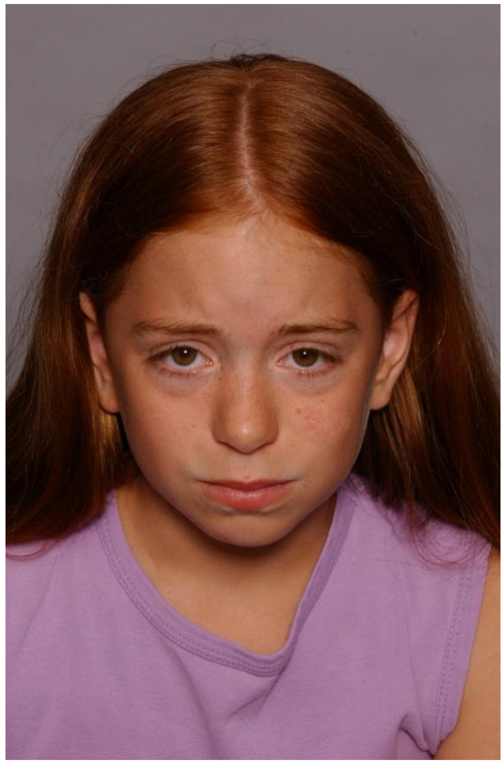
B) Happy

C) Angry

D) Afraid

E) Neutral

Item 31:



A) Sad*

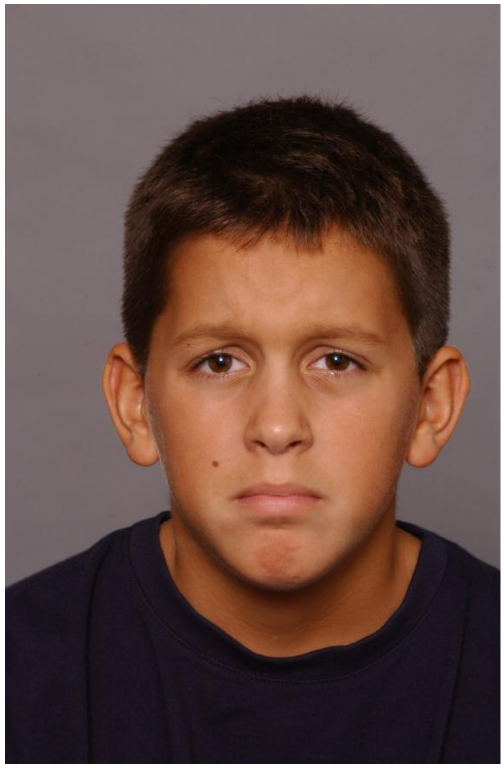
B) Happy

C) Angry

D) Afraid

E) Neutral

Item 32:



A) Sad*

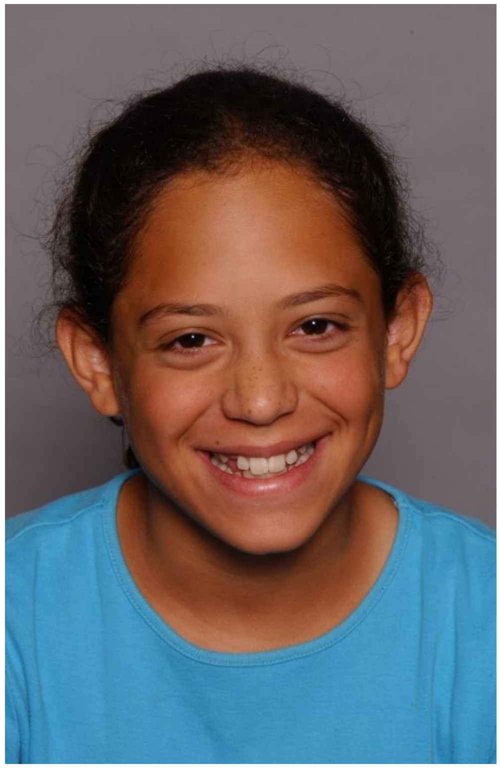
B) Happy

C) Angry

D) Afraid

E) Neutral

Item 33:



A) Sad

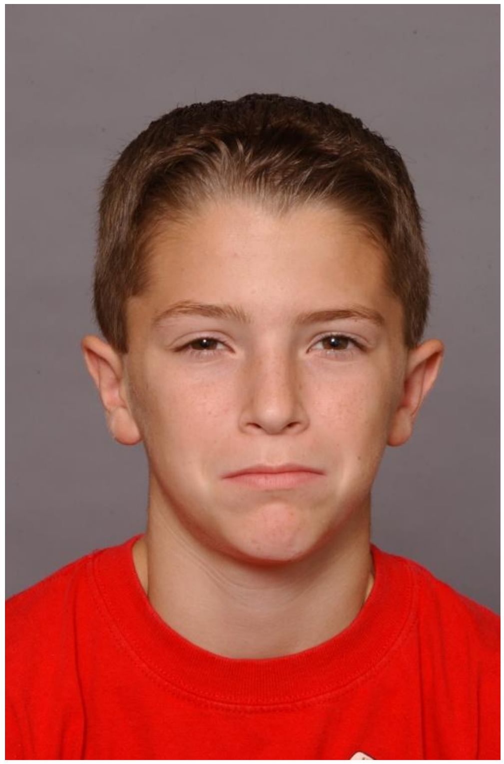
B) Happy*

C) Angry

D) Afraid

E) Neutral

Item 34:



A) Sad*

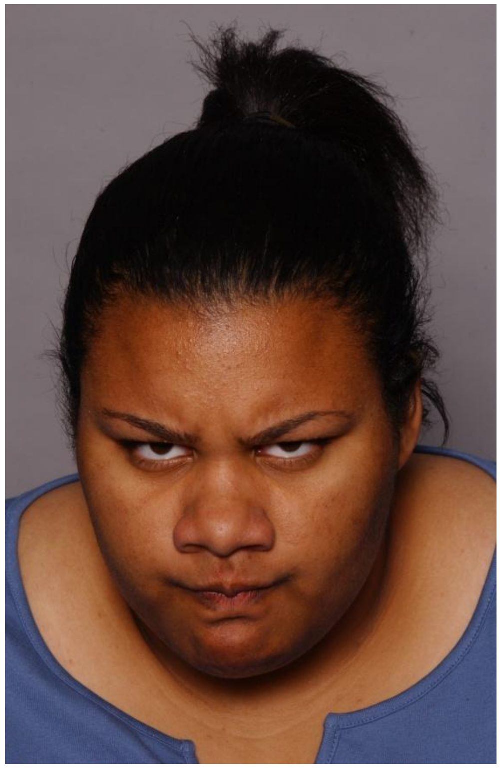
B) Happy

C) Angry

D) Afraid

E) Neutral

Item 35:



A) Sad

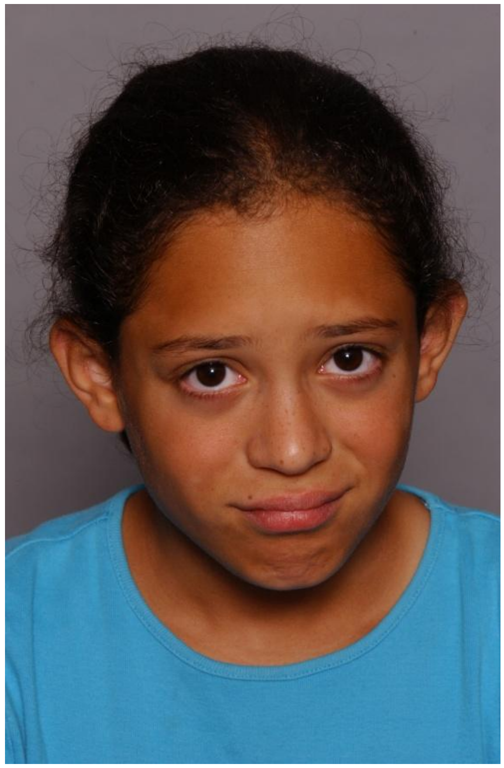
B) Happy

C) Angry*

D) Afraid

E) Neutral

Item 36:



A) Sad*

B) Happy

C) Angry

D) Afraid

E) Neutral

Item 37:

A) Sad

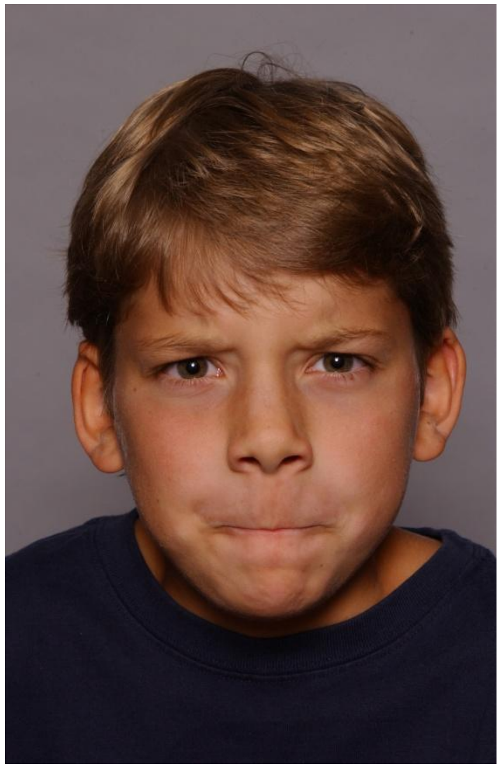
B) Happy

C) Angry

D) Afraid

E) Neutral*

Item 38:



A) Sad

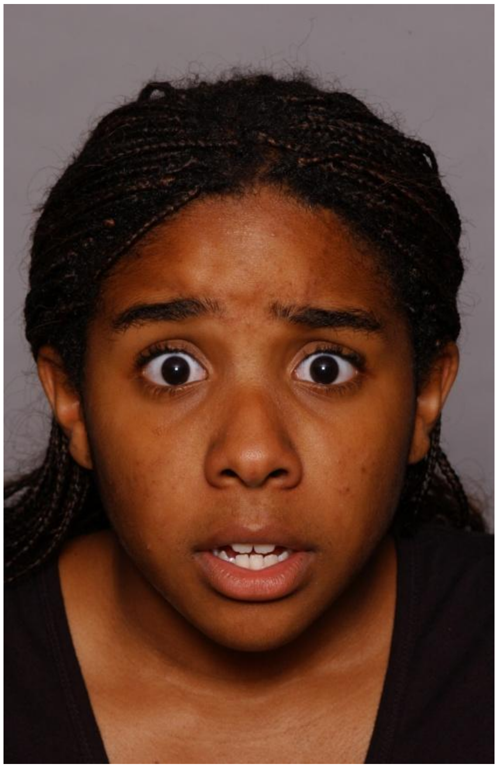
B) Happy

C) Angry*

D) Afraid

E) Neutral

Item 39:



A) Sad

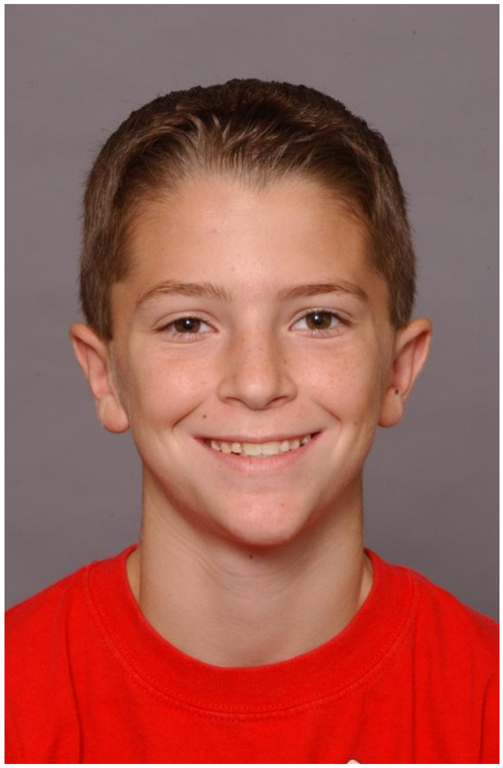
B) Happy

C) Angry

D) Afraid*

E) Neutral

Item 40:



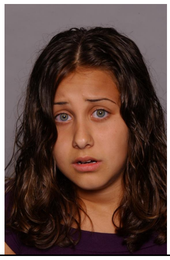
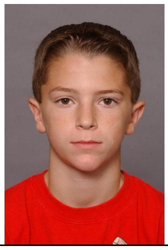
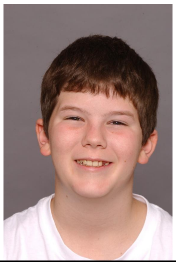
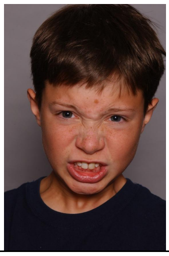
A) Sad

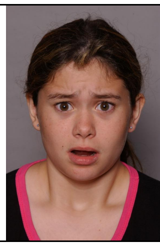
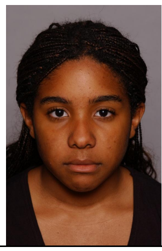
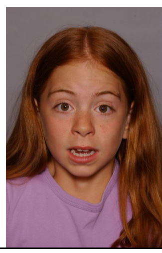
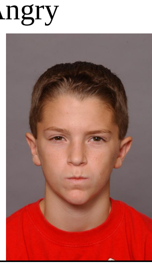
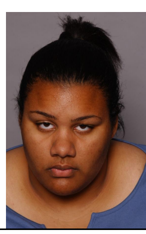
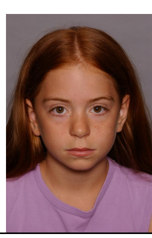
B) Happy*

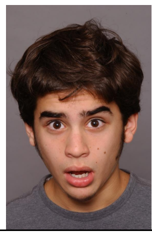
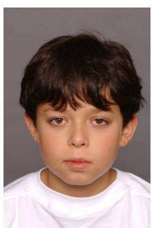
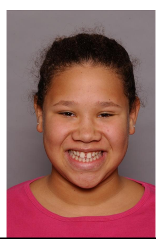
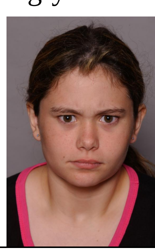
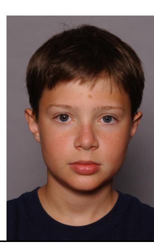
C) Angry

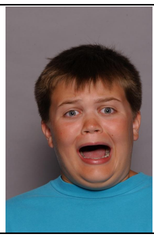
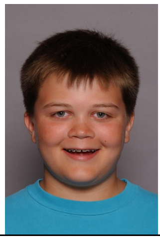
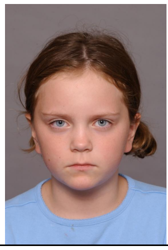
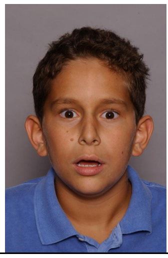
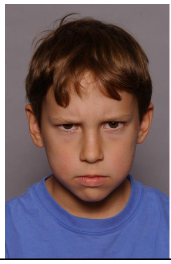
D) Afraid

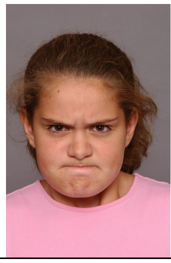
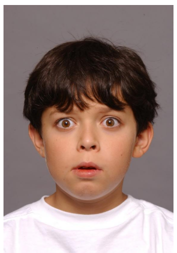
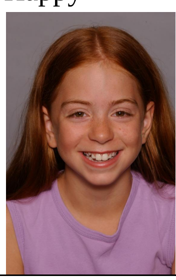
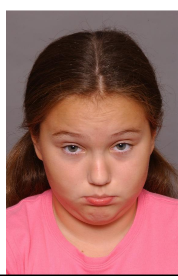
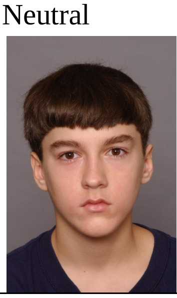
E) Neutral

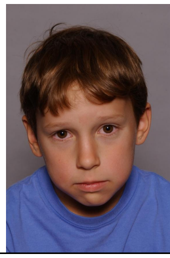
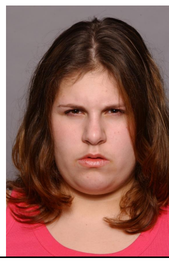
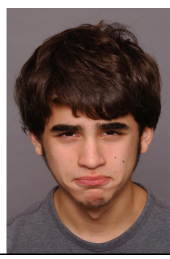
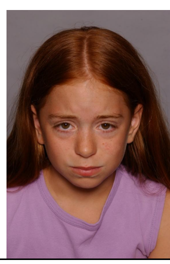
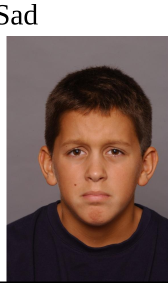
Item	NIMH Picture Set % Agreement	Entire Sample % Correct	Clinical ADHD Group % Correct	TD Group % Correct
1) Afraid 	65	63.0	59.0	66.7
2) Neutral 	100	95.1	94.9	95.2
3) Happy 	100	98.8	100	97.6
4) Angry 	95	97.5	97.4	97.6
5) Happy 	100	93.8	92.3	95.2
6) Afraid	95	92.6	89.7	95.2

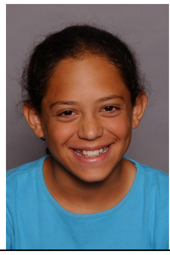
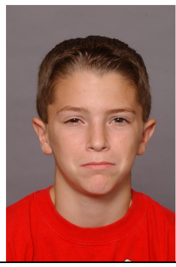
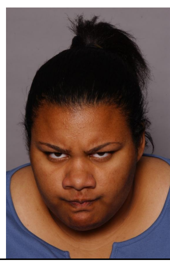
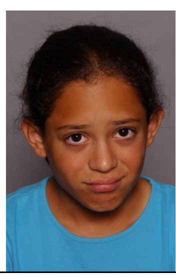
				
7) Neutral 	95	92.6	89.7	95.2
8) Afraid 	95	92.6	89.7	95.2
9) Angry 	100	81.5	82.1	81.0
10) Sad 	15	16.0	7.7	23.8
11) Neutral 	75	93.8	92.3	95.2
12) Happy	100	96.3	94.9	97.6

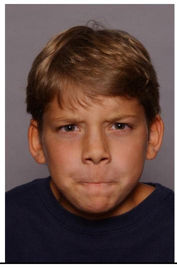
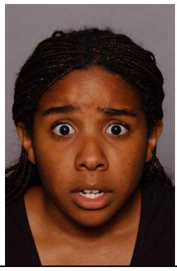
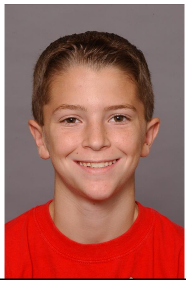
				
13) Afraid 	100	95.1	97.4	92.9
14) Neutral 	100	91.4	89.7	92.9
15) Happy 	95	97.5	97.4	97.6
16) Angry 	80	55.6	66.7	45.2
17) Neutral 	100	91.4	89.7	92.9
18) Afraid	100	95.1	94.9	95.2

				
19) Happy 	100	96.3	94.9	97.6
20) Neutral 	70	53.1	43.6	61.9
21) Afraid 	100	92.6	94.9	90.5
22) Angry 	100	91.4	94.9	88.1

23) Angry 	100	95.1	100	90.5
24) Afraid 	100	92.6	92.3	92.9
25) Happy 	100	97.5	94.9	100
26) Sad 	85	66.7	64.1	69
27) Neutral 	90	98.8	97.4	100

28) Sad 	75	64.2	61.5	66.7
29) Angry 	100	76.5	76.9	76.2
30) Sad 	95	95.1	92.3	97.6
31) Sad 	95	80.2	79.5	81
32) Sad 	100	65.4	64.1	66.7

33) Happy 	100	95.1	94.9	95.2
34) Sad 	80	40.7	48.7	33.3
35) Angry 	90	92.6	100	85.7
36) Sad 	75	82.7	89.7	76.2
37) Neutral 	95	98.8	100	97.6

38) Angry 	90	86.4	94.9	78.6
39) Afraid 	100	95.1	97.4	92.9
40) Happy 	20	100	100	100

Appendix D: Modified Child Social Situation Questionnaire and Percent Correct by Item and Group

Child Social Situations Questionnaire (MALE Version)

Child Social Situations Questionnaire (MALE)

Demonstration question:

A) Pretend that you really like candy and your best friend Sam gives you a whole bag of candy.

i) **What happened in the story?** (Prompt the participant to remember to tell me everything from beginning to end.)

After child's response, ask:

Do you think that something good/positive happened in the story? (Y/N)
How confident/sure are you?
Do you think that something bad/negative happened in the story? (Y/N)
How confident/sure are you?

Repeat above steps with question "B" if child requires additional practice

B) Pretend that you don't like eating Broccoli and your friend Pat gives you Broccoli to eat instead of giving you candy.

Begin the questionnaire by saying "**Now I'm going to read some other stories to you followed by the same type of questions. Let me know what your answers are and I will be writing them down on these pages. Let's begin**"

***Note on Scoring:**

If the story was an *ambiguous positive* situation:

- the Y/N response to whether something good/positive happened in the story was used to score the correctness of the answer (i.e. if participant indicated Yes, then they would get a point for correctly answering the question)
- the confidence rating for whether something good/positive happened was used to calculate the performance calibration measures

If the story was an *ambiguous negative* situation:

- the Y/N response to whether something bad/negative happened in the story would be used to score the correctness of the answer (i.e. if participant indicated Yes, then they would get a point for correctly answering the question)
- the confidence rating for whether something bad/negative happened would be used to calculate the performance calibration measures

The child is asked about whether something positive and whether something negative happened in the story as a foil so as to not direct the child to the valence of the situation.

1) Pretend that you can't find your favorite pack of markers. The last time you remember seeing it was when you were working with a group of other kids on a project. Later that day you see John holding your markers and looking around. John sees you and hands you your markers.
    a. **What happened in the story?**

After child's response, ask:

Do you think that something good/positive happened in the story? (Y/N)

How confident/sure are you?

Do you think that something bad/negative happened in the story? (Y/N)

How confident/sure are you?

2) Pretend that you are standing on the playground playing catch with a kid named Todd. You throw the ball to Todd and he catches it. You turn around, and the next thing you realize is that Todd has thrown the ball and hit you in the middle of your back. The ball hits you hard and it hurts a lot.
    a. **What happened in the story?**

After child's response, ask:

Do you think that something good/positive happened in the story? (Y/N)

How confident/sure are you?

Do you think that something bad/negative happened in the story? (Y/N)

How confident/sure are you?

3) Pretend that you have just arrived at school and you were really in a rush. When you get to the classroom you realize you left one of your books in the hallway. You go into the hallway to get it and see a kid named Brandon looking through your book. Brandon sees you and hands you your book.
    a. **What happened in the story?**

After child's response, ask:

Do you think that something good/positive happened in the story? (Y/N)

How confident/sure are you?

Do you think that something bad/negative happened in the story? (Y/N)

How confident/sure are you?

4) Pretend that you are walking down the hallway in school. You're carrying your books in your arm and talking to a friend. Suddenly, a kid named Brett bumps you from behind. You stumble and fall and your books go flying across the floor. The other kids in the hall start laughing.
    a. **What happened in the story?**

After child's response, ask:

Do you think that something good/positive happened in the story? (Y/N)

How confident/sure are you?

Do you think that something bad/negative happened in the story? (Y/N)

How confident/sure are you?

5) Pretend that you are walking to school and you're wearing brand new sneakers. You really like your new sneakers and this is the first day you have worn them. Suddenly, you are bumped from behind by a kid named Alex. You stumble into a mud puddle and your new sneakers get muddy.

a. **What happened in the story?**

After child's response, ask:

Do you think that something good/positive happened in the story? (Y/N)

How confident/sure are you?

Do you think that something bad/negative happened in the story? (Y/N)

How confident/sure are you?

6) Pretend that you and your classmates are about to play a game of baseball. Patrick and Steven are chosen as team captains. Pretend that every time Patrick has been captain he has always chosen you last. You see Patrick looking at you before he begins to choose kids. Patrick chooses you close to first.
    a. **What happened in the story?**

After child's response, ask:

Do you think that something good/positive happened in the story? (Y/N)

How confident/sure are you?

Do you think that something bad/negative happened in the story? (Y/N)

How confident/sure are you?

7) Pretend that you and your class went on a field trip to the zoo. You stop to buy a coke. Suddenly, a kid named David bumps your arm and spills your coke all over your shirt. The coke is cold, and your shirt is all wet.
    a. **What happened in the story?**

After child's response, ask:

Do you think that something good/positive happened in the story? (Y/N)

How confident/sure are you?

Do you think that something bad/negative happened in the story? (Y/N)

How confident/sure are you?

8) Pretend that you can't find your favorite pencil but you are sure that you brought it to school with you. Later that day you see one of the kids in your class named Jesse walking towards your desk with the pencil. Jesse puts your favorite pencil on your desk.
    a. **What happened in the story?**

After child's response, ask:

Do you think that something good/positive happened in the story? (Y/N)

How confident/sure are you?

Do you think that something bad/negative happened in the story? (Y/N)

How confident/sure are you?

Item	Entire Sample % Correct	Clinical ADHD Group % Correct	TD Group % Correct
1) Pretend that you can't find your favorite pack of markers. The last time you remember seeing it was when you were working with a group of other kids on a project. Later that day you see John holding your markers and looking around. John sees you and hands you your markers.	92.6	94.9	90.5
2) Pretend that you are standing on the playground playing catch with a kid named Todd. You throw the ball to Todd and he catches it. You turn around, and the next thing you realize is that Todd has thrown the ball and hit you in the middle of your back. The ball hits you hard and it hurts a lot.	86.4	92.3	81.0
3) Pretend that you have just arrived at school and you were really in a rush. When you get to the classroom you realize you left one of your books in the hallway. You go into the hallway to get it and see a kid named Brandon looking through your book. Brandon sees you and hands you your book.	79	82.1	76.2
4) Pretend that you are walking down the hallway in school. You're carrying your books in your arm and talking to a friend. Suddenly, a kid named Brett bumps you from behind. You stumble and fall and your books go flying across the floor. The other kids in the hall start laughing.	97.5	94.9	100
5) Pretend that you are walking to school and you're wearing brand new sneakers. You really like your new sneakers and this is the first day you have worn them. Suddenly, you are bumped from behind by a kid named Alex. You stumble into a mud puddle and your new sneakers get muddy.	91.4	92.3	90.5
6) Pretend that you and your classmates are about to play a game of baseball. Patrick and Steven are chosen as team captains. Pretend that every time Patrick has been captain he has always chosen you last. You see Patrick looking at you before he begins to choose kids. Patrick chooses you close to first.	95.1	94.9	95.2
7) Pretend that you and your class went on a field trip to the zoo. You stop to buy a coke. Suddenly, a kid named David bumps your arm and spills your coke all over your shirt. The coke is cold, and your shirt is all wet.	92.6	89.7	95.2
8) Pretend that you can't find your favorite pencil but you are sure that you brought it to school with you. Later that day you see one of the kids in your class named Jesse walking towards your desk with the pencil. Jesse puts your favorite pencil on your desk.	96.3	94.9	97.6

Appendix E: Confidence Rating Scale for GKT, ERT, and CSSQ

Confidence Rating Scale

0 1 2 3 4 5 6 7 8 9 10

0	5	10
I'm Not Sure if I'm right	Good Chance I'm Right	I'm Definitely Right

Appendix F: Cognitive Effort Rating Scales

Cognitive Effort

1) How hard did you find this task?

(Please mark on the picture to indicate)

2) How hard did you try on this task?

(Please circle the number to indicate)

1 2 3 4 5

I did not try at all

I tried very hard

Appendix G: Demographics Form

Demographic Questionnaire

- **This form asks for information about your child**
- **Please indicate beside a question if you prefer not to answer**

1. Child's Date of Birth:
   Day ___ Month ___ Year _______

2. Date form was completed: ________/__________/_____________
   (dd / mm / yyyy)

3. This form is being completed by (check those that apply):
   a. ☐ Biological mother
   b. ☐ Biological father
   c. ☐ Step-mother
   d. ☐ Step-father
   e. ☐ Adoptive mother
   f. ☐ Adoptive father
   g. ☐ Biological grandparent
   h. ☐ Foster parent
   i. ☐ Child Welfare Worker
   j. ☐ Other: ____

4. Child's gender: ☐ Male ☐ Female

5. Child's Country of Birth: ______________________________

6. If not born in Canada, child's date of immigration: Month ______ Year _______

7. The following categories are used by Statistics Canada to understand the ethnic and cultural background of Canadians.

To which ethnic or cultural groups(s) did your child's ancestors belong? Please specify as many groups as applicable.

a. ☐ British Isles origins
b. ☐ French origins
c. ☐ Aboriginal origins
d. ☐ North American origins
e. ☐ Caribbean origins
f. ☐ Latin, Central and South American origins
g. ☐ European origins
h. ☐ Western European origins
i. ☐ Northern European origins
j. ☐ Scandinavian origins
k. ☐ Eastern European origins
l. ☐ Baltic origins
m. ☐ Czech and Slovak origins
n. ☐ Southern European origins
o. ☐ Balkan origins
p. ☐ African origins
q. ☐ Arab origins
r. ☐ Maghrebi origins
s. ☐ West Asian origins
t. ☐ South Asian origins
u. ☐ East and Southeast Asian origins
v. ☐ Indo-Chinese origins
w. ☐ Oceania origins
x. ☐ Pacific Islands origins

8. Caregiver Information

	Parent 1		**Parent 2**	
Relationship to child				
Country of Birth				
Current marital status				
Age				
Highest Level of Education	1. Up to Grade 8		1. Up to Grade 8	
	2. Grade 9 – 12		2. Grade 9 – 12	
	3. Some post secondary		3. Some post secondary	
	4. Completed post secondary		4. Completed post secondary	
	5. Do not know		5. Do not know	
Language(s) spoken at home				
Confidence with reading / writing English:	☐ Low ☐ Moderate ☐ High		☐ Low ☐ Moderate ☐ High	
Confidence with understanding spoken English:	☐ Low ☐ Moderate ☐ High		☐ Low ☐ Moderate ☐ High	
Occupation *****Please be as specific as possible				

9. **<u>Currently</u>**, is your child taking any medicine that is <u>not antibiotic</u>?

a. ☐ No
b. ☐ Yes:
1) (i) Name of medication ______________________(ii) Dose _____________
(iii) For what reason?________________(iv) When started________________

2) (i) Name of medication ______________________(ii) Dose _____________
(iii) For what reason?________________(iv) When started________________

3) (i) Name of medication ______________________(ii) Dose _____________
(iii) For what reason?________________(iv) When started________________

c. ☐ Don't know

10. **<u>In the past</u>**, has your child taken medicine that is <u>not antibiotic</u>?

a. ☐ No
b. ☐ Yes:
1) (i) Name of medication ______________________(ii) Dose _____________
(iii) For what reason?________________(iv) When started________________

2) (i) Name of medication ______________________(ii) Dose _____________
(iii) For what reason?________________(iv) When started________________

3) (i) Name of medication ______________________(ii) Dose _____________
(iii) For what reason?________________(iv) When started________________

c. ☐ Don't know

This study is about metacognition, or how well your son/daughter can monitor his/her performance on tasks. How would you rate your child's metacognition on tasks?

1. My son/daughter has a good sense of whether his/her work is completed correctly or accurately.

   1 2 3 4 5 6 7

   No, poor tracking of accuracy of work — Yes, excellent tracking of accuracy of work

2. Do you have any comments about your son/daughter's performance monitoring on tasks? Be specific and/or give examples.

   ____
   ____
   ____
   ____
   ____
   ____

Thank you for completing this form.