CAPORASO, JESSICA S., Ph.D. Revisiting the Association Between Executive Function and Social Competence: A Process-Oriented Account of Social and Moral Development. (2021) Directed by Dr. Stuart Marcovitch. 115 pp.

The goal of the current dissertation study was to examine the longitudinal association between executive function (EF) and social competence (SC). Previous research has shown that SC in peer conflict situations is an important skill and that EF supports the development of SC during preschool. However, less is known about the mechanism by which EF supports SC development. In addition, there is evidence to suggest that not all children require the same level of EF to behave in a competent manner. The current study explored the possibility that children's temperament or moral reasoning abilities differentiated between these children. The secondary goal of the study was to examine how EF and classroom behavior contributed to children's moral development over the preschool year.

To address these goals, 86 4- to 5-year-old children were tested at the beginning of the preschool year (T1) and 33 of these children (because of retention difficulties associated with COVID-19) were tested during the summer following the preschool year (T2). Participants completed a social problem solving task, a moral reasoning interview, and an EF task battery at both time points. Parents completed a temperament questionnaire at T1 and teachers reported on children's classroom behavior at T2. Results revealed that T1 EF predicted SC, but only for participants high in temperamental surgency. T1 EF additionally predicted T2 moral reasoning, but in opposite directions depending on participants' classroom prosocial behavior. The results broadly provide support for current theories regarding the development of both SC and moral reasoning during the preschool years.

REVISITING THE ASSOCIATION BETWEEN EXECUTIVE FUNCTION AND SOCIAL COMPETENCE: A PROCESS-ORIENTED ACCOUNT OF SOCIAL AND MORAL DEVELOPMENT

by

Jessica S. Caporaso

A Dissertation

Submitted to

the Faculty of The Graduate School at

The University of North Carolina at Greensboro

in Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

Greensboro

2021

Approved by

Dr. Stuart Marcovitch

Committee Chair

APPROVAL PAGE

This dissertation written by Jessica S. Caporaso has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Chair	Dr. Stuart Marcovitch
Committee Members	Dr. Janet Boseovski
	Dr. Julia Mendez Smith
	Dr. Rosemery Nelson-Gray

4/29/2021

Date of Acceptance by Committee

4/29/2021

Date of Final Oral Examination

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Stuart Marcovitch. I will always marvel at the fact that I found the perfect fit in an advisor with only applying to two local schools. I offer my deepest gratitude to him for allowing me to flourish as a researcher, for giving me incredible opportunities, and for offering unwavering support and guidance throughout. It has meant so much to me and I will never be able to thank him enough. I will always treasure our many intellectually stimulating debates, inside jokes, shared (bad) tastes in tv and food, and heartfelt conversations.

I would also like to thank my committee members for their invaluable feedback and support, starting with Janet Boseovski. I said in her promotion letter that I credit her for the strength of my research program and that is the absolute truth. I thank her for always pushing me to think critically, for her input when I fell short of doing so myself, and for the times that she has given me the much-needed confidence boost. I have been so lucky to be a part of a lab with not just one, but two incredible advisors. Thank you to Julie Mendez Smith, for teaching me how to fully appreciate the complexity of development (I still refer back to discussions from her classes) and showing me that our work does indeed matter. I will always admire and be inspired by her commitment to equity, allyship, and using her work to better the lives of children in our community. Thank you to Rosemery Nelson-Gray, whom I was lucky enough to have assigned to my committee. I thank her for her constant words of encouragement, her smiles and greetings that I have sorely missed during the pandemic, and her commitment to her students, this department, and women in psychology.

Thank you to the 12 childcare centers – the directors and the preschool teachers – for participating in my dissertation study. Their eagerness to support my research was touching and I truly would not have been able to complete this study without it.

To my support system throughout this process, particularly Kimmy Marble and Kathleen Bettencourt. Thank you for being constant sources of comfort, advice, encouragement, and friendship. I cannot put into words how truly grateful I am for them and how much I value my relationships with them. To Chris Erb and Courtney Ball, for being my mentors, my cheerleaders, and for providing me with the resources I needed to strengthen my program of work. Thank you to the many research assistants that assisted me with my work, but particularly Destiny Langley. Her help with dissertation data collection, her willingness to do whatever needed to be done, and her never-ceasing positive attitude were incredibly appreciated. I would also like to thank Luna, for being the best dog in the entire world, as well as Andrew McMahon, Harry Potter (but explicitly not she-who-must-not-be-named), and Doctor Who for providing comfort and companionship during the long nights of writing.

Finally, I would like to thank my family, for none of this would be possible without their love, guidance, and unconditional support- my parents, Vickie and Chris Stark, my in-laws, Carolyn and Joe Caporaso, my many brothers and sisters (but special shoutout to Kristen for being my live-in babysitter for a year and providing me with Maxie B's whenever work deemed it necessary), and my grandparents and Aunt Nancy. Thank you for being there to celebrate my successes, listen during my struggles, and cheering me on every step of the way.

Most importantly, thank you to Joe, my partner in life, and Jac, my source of endless joy. Joe, thank you for *everything*, from the numerous cups of early morning coffee and late night tea, to picking up my slack around the house when I was overwhelmed by work, and for always being there throughout for comfort, laughter, and love. Jac, thank you for being the sweetest, cutest, hilarious, and absolute best son (and for eventually being a champion sleeper, allowing me to get work done after your 7:30 p.m. sharp bedtime). Even on the worst days, the thought of seeing you at the end of the day always made everything a thousand times better. Everything I have done has been for you and your dad. As Andrew says, "I'll never stop trying, you're my silver lining."

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I: INTRODUCTION

Peer conflict situations are common in the preschool classroom (Chen et al., 2001; Raikes et al., 2013; Spivak, 2016). A major milestone of social competence (SC) development is the ability to resolve these conflicts in an appropriate, non-aggressive, manner (e.g., Denham et al., 1994; Dodge et al., 1986). In fact, Rubin and Rose-Krasnor (1992) argued that interpersonal problem solving is the very definition of SC in children. Preschool children who respond to conflict in an aggressive manner, also referred to as reactive aggression, are more likely to be rejected by their peers (Sebanc, 2003) and have adjustment issues during the transition from preschool to kindergarten (Denham et al., 2013), possibly because of a distinct pattern of maladaptive learning behaviors (Coolahan et al., 2000).

By the end of preschool, most children have learned how to respond to conflict without aggression (e.g., Alink et al., 2006; Carbonneau et al., 2016; Hill et al., 2006; NICHD Early Child Care Research Network, 2004), which suggests that some level of preschool aggression is normative, but a small number of children continue to display high amounts of aggression throughout childhood (e.g., Hill et al., 2006; NICHD Early Child Care Research Network, 2004). These children are at high risk for later adjustment difficulties, including poor academic, social, and legal outcomes (e.g., Brook & Newcomb, 1995; Moffitt et al., 2001; NICHD Early Child Care Research Network, 2004). Thus, it is important to determine the factors that support the decrease of normative aggression in preschool and the differences between children who are aggressive past preschool and those that are not aggressive past preschool.

Executive function (EF), the cognitive control of thoughts, emotions, and behaviors (Zelazo & Carlson, 2012), has been identified as an important factor in the development of SC. Previous work has identified three components of EF: (a) working memory, or the ability to hold and manipulate information, (b) response inhibition, or the ability to suppress a prepotent response, and (c) cognitive flexibility, or the ability to shift to a new response (e.g., Diamond, 2013; Lehto

et al., 2003; Miyake et al., 2000). A review of existing research reveals a robust association between EF and aggression, and reactive aggression in particular (e.g., Denham et al., 2014; Poland et al., 2016; Rathert et al., 2011). This association has been found in atypical samples (e.g., Cole et al., 1993; Razza & Blair, 2009; Schoemaker et al., 2014), as well as in typically developing samples (e.g., Ciairano et al., 2006; Denham et al., 2014; Kochanska & Knaack, 2003), across early childhood (e.g., Hughes et al., 2000; Poland et al., 2016) and through late childhood (e.g., Granvald & Marciszko, 2016; McQuade et al., 2013). The EF–SC association is typically observed cross-sectionally, but it has also been demonstrated that early EF predicts later SC (e.g., Kochanska & Knaack, 2003; Nigg et al., 1999; O'Toole et al., 2019).

In addition, the EF–SC association is observed across a variety of methods used to measure both EF and SC. Although some studies use parent questionnaires to assess EF (e.g., White et al., 2013), most studies use lab-based tasks to measure EF but vary in which aspects of EF they assess. The EF-SC association has been observed when one aspect of EF was measured (e.g., response inhibition, Utendale et al., 2011), when an aggregate score across measures of all EF components was used (e.g., Alduncin et al., 2014), and when the different components of EF were measured separately (e.g., Caporaso et al., 2019; Granvald & Marciszko, 2016; O'Toole et al., 2019). Assessments of SC also vary; some studies used parent or teacher questionnaires to assess children's SC (e.g., Rathert et al., 2011), while other studies used observational methods (e.g., Olson et al., 2011), and yet others assessed children's responses to hypothetical situations (e.g., Caporaso et al., 2019; Denham et al., 2014). In sum, a large body of empirical evidence suggests that EF is a vital skill in the promotion of SC and the hindrance of aggression.

Caporaso et al. (2019) linked the rapid development of EF during the preschool years (Zelazo et al., 2013) to the age-related changes in SC that are also observed during preschool (e.g., Alink et al., 2006; Carbonneau et al., 2016; Hill et al., 2006), but several questions remain regarding this association. First, the mechanism by which EF is related to SC remains unclear. Is it the case that EF is necessary for the development of SC over time or is EF only needed in the moment of conflict? Second, it is possible that some children do not need to rely on EF to refrain from aggression. Indeed, there is a substantial group of children who are never or rarely aggressive, even during the tumultuous toddler years (e.g., Hill et al., 2006; Tremblay et al., 2004). It seems

unlikely that these children experience the same EF demands to control an aggressive response compared to children on the other trajectories, yet the potential moderators of the EF-SC relation have not been explored. Thus, the primary aim of the current study was to examine the EF mechanisms that support SC development as well as address the possibility that EF is only important for some children in conflict situations through the use of a short-term longitudinal design.

Mechanisms of EF in Relation to SC

The most common explanation for the role of EF in SC is that competent behaviors are a product of controlled, effortful processes that prevent the action of prepotent incompetent tendencies (e.g., Beauchamp & Anderson, 2010; DeWall et al., 2011; Landry et al., 2009; Poland et al., 2016; Rhoades et al., 2009; Riggs et al., 2006; Walker & Henderson, 2012; White et al., 2013). For example, Denham et al. (2014) suggested that EF helps children "inhibit prepotent responses while activating alternative, subdominant responses" (p. 184). This explanation suggests that EF provides in the moment support for SC, but it does little to explain how EF might support the development of SC over the course of the preschool years. Gradual increases in EF ability may relate to gradual increases in children's ability to respond to conflict without aggression. That is, EF continually provides in the moment support for competent behaviors at its current developmental level and eventually EF abilities become strong enough to override dominant aggressive responses consistently. However, it is also possible that EF is necessary for the emergence of SC at a deeper, more conceptual level.

This notion is similar to the *expression* and *emergence* accounts of the well-documented association between EF and theory of mind, or children's understanding of others' mental states (Moses, 2001). The *expression* account posits that theory of mind and EF are related because of the EF demands placed on children while completing theory of mind tasks. Conversely, the *emergence* account posits that EF is necessary for the conceptual development of theory of mind. Without EF, children may be unable to remove themselves from their own perspective to consider the perspectives of others. As EF develops, children gain the reflective capacity to do so and can begin to construct sophisticated cognitive concepts about the mental states of others.

EF may play a similar role in the development of SC. Tremblay (2010) suggested that self-regulatory abilities may be the mechanism by which children replace the aggressive tendencies that are characteristic of toddlerhood with competent response options as they approach middle childhood. Tremblay argues that self-regulatory abilities (i.e., EF) not only override dominant aggressive responses but actively replace those responses with a dominant competent response. Once children have a dominant competent response, children may not need EF to the same degree to act in a competent manner. But how exactly does EF contribute to the replacement of aggressive responses with more appropriate responses? Perhaps early competency in EF allows children to practice appropriate responses during conflict situations frequently enough for them to become the viable, preferable options in provocation situations (Eisenberg et al., 1995; O'Toole et al., 2019). Like the emergence account of theory of mind, this process would indicate that EF assists children with the development of a SC concept.

The importance of having a developed SC concept was outlined by the social information processing (SIP) model (Crick & Dodge, 1994). The SIP model suggests that social behavior is influenced by preexisting knowledge of social interactions, stored as latent mental structures in the "database". The database contains scripts and schemas for social interaction, as well as internal working models of relationships, a concept adapted from the attachment literature (e.g., Bowlby, 1982). These representations are constructed through continued social interaction; children create memories of social experiences, which are then organized and stored in the form of scripts and schemas. As children experience provocation, the response options that are most likely to come to mind are those that have been frequently used in the past and have the most enduring mental representations. Children then consider their self-efficacious beliefs and choose to enact a response that best matches their self-assessed social skills. Thus, having the chance to practice competent responses may be a critical component in the development of SC and EF is a key ability that helps children have such opportunities.

An emergence account of EF in SC development would potentially explain why the EF-SC association is observed across multiple different types of measurement, particularly measurements of children's responses to hypothetical situations. The hypothetical situation tasks remove participants from the typical context of peer provocation and likely evoke fewer negative

emotions. Because of this, these tasks should have lower EF demands than real-life provocation experiences (and by proxy, the measures that capture children's behaviors in these experiences). Indeed, Carlson et al. (2015) found that EF was related to theory of mind tasks that objectively had less EF demand and reasoned that this provides evidence that EF supports the development of a theory of mind concept (i.e., the emergence account).

In addition, the finding that early EF predicts later theory of mind but early theory of mind seldom predicts later EF is often cited as evidence of the emergence account (Carlson et al., 2004; Marcovitch et al., 2015; Müller et al., 2012). This pattern of results provides support for the emergence account because it suggests that sufficient EF ability precedes the development of theory of mind. If it were simply an issue of task demand, it is expected that early EF would not predict later theory of mind (directly or indirectly) after controlling for current EF ability. The emergence account would also not be supported if there was a symmetrical relation between theory of mind and EF (i.e., early EF predicts later theory of mind and early theory of mind predicts later EF; Carlson et al., 2015).

Early EF does predict later competence (e.g., Kochanska & Knaack, 2003; Nigg et al., 1999; O'Toole et al., 2019), which provides some support for an emergence account of EF and SC development. However, these studies did not control for current EF ability or test if early SC predicts later EF. There is a possibility that early SC could predict EF development, perhaps because early social conflict provides children with the opportunity to practice EF skills (Landry & Smith, 2010) or because the general ambiguity of conflict situations could spur EF development to meet the demands of children's social contexts (Hala et al., 2010). The current study addressed these gaps in the literature by including later EF in the model and testing the possibility for a symmetrical relation between EF and SC.

Indicators of Individual Differences in EF Demands During Peer Provocation

The application of the expression and emergence accounts to SC provides a promising avenue to further our understanding of the EF-SC association. However, SC differs from theory of mind and these differences need to be taken into consideration. Most children develop theory of mind

skills on a similar trajectory as their EF develops during the preschool years. Yet, it is estimated that between 16-28% of children never or rarely show displays of aggression, even as very young children with limited EF capabilities (Hill et al., 2006; Tremblay et al., 2004). It stands to reason that EF is not necessary for either the expression *or* the emergence of SC for these children and that there are other factors present in early childhood that may predict a child's level of SC.

There is also evidence to suggest that the EF demands in peer provocation situations differ for children based on other social cognitive factors. Ellis et al. (2009) found that EF was only related to aggressive behavior among 9- to 11-year-old boys who were more likely to attribute hostile intent to the transgressor in ambiguous conflict situations. They concluded that boys who attributed benign intent did not need control in conflict situations because they likely lacked a dominant aggressive response, as such a response is more characteristic of children who attribute hostile intent to transgressors. In a sample of 4- to 5-year-old boys and girls, Caporaso and Marcovitch (2017) found that participants with low EF who favorably rated acts of reactive aggression chose significantly more aggressive responses to conflict than participants with equally low EF who unfavorably rated the same acts. Similar to the conclusion drawn by Ellis et al., perhaps children that rated the aggressive acts unfavorably were less likely to have a dominant aggressive response than the children who rated aggressive responses favorably. Both studies illustrate that children's knowledge and understanding of sociomoral norms may reflect differences in children's prepotent responses to conflict, and thus indicate differential needs for control in difficult social situations.

The current study focuses on two potential factors that could affect the level of EF required for SC in conflict situations: temperament and moral reasoning. Temperament refers to genetically-based individual differences in children's reactivity and regulatory abilities (Rothbart & Bates, 2006). Temperamental differences are evident early in infancy and persist throughout the lifespan. Moral reasoning captures children's knowledge of sociomoral rules, specifically the understanding that aggression is wrong across all contexts because of the harm it causes another individual (Smetana et al., 2014). Individual differences in children's moral reasoning, particularly regarding reactive aggression, are observed during the preschool years (Caporaso et

al., 2021). Both factors may differentiate children who do and do not need to rely on EF for the expression *or* emergence of SC during the preschool years.

TEMPERAMENT

Temperament is composed of three dimensions: surgency, negative affect, and effortful control (Rothbart & Bates, 2006). Surgency refers to children's activity level, level of sensation-seeking, extraversion, and general avoidance and approach tendencies. Negative affect refers to the frequency and intensity of children's experiences with negative emotions. Effortful control refers to children's inhibitory and attentional control, as well as children's enjoyment of low-intensity activities. Because effortful control is primarily centered on children's behavioral regulation, it is a construct that highly overlaps with EF (see Zhou et al., 2012, for review). In fact, lab-based tasks of effortful control are isomorphic to lab-based tasks for the response inhibition component of EF (Carlson, 2005; Garon et al., 2008; Kochanska et al., 1997; Kochanska et al., 1996). Therefore, it is no surprise that a relation between effortful control and SC is well-documented (e.g., Eisenberg et al., 1994; Eisenberg et al., 2003; Kochanska & Knaack, 2003; Olson et al., 2005; Rathert et al., 2011). Because of the theoretical and empirical overlap between EF and effortful control, the current study focused primarily on surgency and negative affect.

There is a noted association between negative affect and aggression (e.g., Eisenberg et al., 2009; Cohen & Mendez, 2009; Gilliom et al., 2002; Mendez et al., 2002; Moran et al., 2013; Zeman et al., 2002). Negative affect encompasses sadness, fear, and frustration/anger. Anger, in particular, is considered to be an important predictor of reactive aggression (Dollar & Calkins, 2019; Orobio de Castro, 2004; Orobio de Castro et al., 2012; Runions & Keating, 2010) and has been shown to relate to preschool aggression and general SC (Eisenberg et al., 2001). Feelings of anger may motivate children towards aggressive responses and make it difficult to generate competent alternatives (Lemerise & Arsenio, 2000). In addition, intense feelings of anger may interfere with EF processes and make it harder for children to control their behavior in provocation situations (Blair, 2014). Further, children might need a certain level of EF ability to regulate strong feelings of anger (Calkins & Marcovitch 2010; Moran et al., 2013). It is thought

that children who are unable to control their anger may have a harder time developing the social skills necessary to behave competently in a variety of social situations (Cohen & Mendez, 2009; Dollar & Calkins, 2019). Even in circumstances when not directly involved in the provocation situation, children who report that they would feel angry if the situation happened to them are more likely to endorse aggressive response options (Denham et al., 2013). This suggests that anger not only interferes with real-life circumstances of provocation but also the ability to recognize that competent options are the most appropriate responses to provocation.

Sadness has also been linked to SC (Eisenberg et al., 2005; Lemry et al., 2002), perhaps because children prone to sadness may have global social skills deficits and lack the motivation to work on improving these skills (Capaldi, 1992; Eisenberg et al., 2009). It is also theorized that aggressive children may face higher rates of peer rejection, which may lead to more affective displays of sadness, fewer opportunities to practice social skills, and consequently worse SC (Eisenberg et al., 2009; Lansford et al., 2010). However, it seems that the pathways in which sadness would relate to aggression would become stronger over time as these cascading processes unfold and may not be evident in early childhood.

Although much of the focus has been on how effortful control and negative affect relate to SC, there is evidence to suggest that high levels of surgency are also linked to problematic social behavior. Children high in surgency, often referred to as “exuberant” children (Stifter et al., 2008), do have marked social advantages because they are more extraverted and less inhibited in social situations. Preschool children high in surgency have been found to display more social interest, positive affect, and goal-directed cooperation during structured play tasks with an unfamiliar peer (Degnan et al., 2011). However, there appears to be a double-edged sword for exuberant children; the lack of social inhibition that lends itself to making friends with unfamiliar peers can also lead to negative peer interactions (Dollar & Stifter, 2012). This could be particularly problematic when surgent children seek to play with peers who do not match their activity level, leading to disruptive play behaviors and conflict with their less-active peers (Mendez et al., 2002). The approach tendencies associated with surgency are also associated with frustration, particularly when prevented from accomplishing a goal or receiving a reward (Rothbart et al., 2000). Consequently, highly surgent preschool children who may be socially

skilled in some domains also exhibit more negative play behaviors with unfamiliar peers (Dollar & Stifter, 2012) and are rated as more aggressive by parents and teachers both cross-sectionally (Gunnar et al., 2003) and longitudinally (Berdan et al., 2008; Stifter et al., 2008).

Overall, children's temperament may be indicative of their baseline tendencies towards aggression, as children high in surgency and negative affect may be more likely to engage in aggression than others. Consequently, these children need to rely more on EF to refrain from aggression. Perhaps children low in surgency and negative affect are the same children who never or rarely show aggression throughout early childhood. These children would not need to rely on EF either for the expression or emergence of SC. Thus, temperament may interact with EF in the development of SC.

Moral Reasoning

Children's moral reasoning abilities have also been studied in the context of aggression and could additionally identify which children require EF in provocation situations. A large body of work from the social domain perspective has determined that children understand that unprovoked aggression is morally wrong at a very early age, meaning that children understand unprovoked aggression as an immoral act independent of rules and authority (e.g., Smetana & Braeges, 1990; Smetana et al., 2014). However, moral reasoning regarding aggression in response to provocation (i.e., reactive aggression or retaliation) shows a more protracted developmental trend and continues to develop through middle childhood. Smetana et al. (2003) found that kindergarten children rated provoked aggression as less serious and less morally wrong than 2nd and 4th grade children. Similarly, Caporaso et al. (2021) found that children's overall negative ratings of reactive aggression increased between 4 and 5 years of age.

One important question that has recently regained interest in the field is how moral reasoning about aggression relates to actual behavior (Jambon & Smetana, 2017). Arsenio and Lemerise (2004; but see also Dodge & Rabiner, 2004) explicitly stated that moral reasoning is an important component of the SIP model database. According to the SIP model (Crick & Dodge, 1994), the database exerts considerable influence on how children solve social problems,

including biasing children towards certain response options and away from others (Burks et al., 1999; Zelli et al., 1999). An extension of the SIP model, the Response and Evaluation Decision (RED) model (Fontaine & Dodge, 2006), further suggests that children who consider social and moral rules prior to the enactment of a response to conflict are more likely to respond competently.

A few studies have examined associations between moral reasoning and aggressive behavior. Jambon and Smetana (2017) found that unprovoked (i.e., proactive) aggression in 4- and 6-year-old children was related to less mature moral reasoning but conversely, reactive aggression was related to more mature moral reasoning. However, the scenarios children were asked to judge were all focused on unprovoked aggression and not retaliatory or reactive aggression. Children who engage in reactive aggression may view retaliatory acts differently than situations of proactive, unprovoked aggression. When asked to reason about retaliatory acts, 7- to 9-year-old children classified as "aggressive" by their peers rated retaliation as more acceptable than non-aggressive children (Gasser et al., 2012). Notably, the two groups of children did not differ on the criterion judgments that are indicative of having a mature moral concept. Further, preschoolers rated highly on measures of reactive aggression display distinct patterns of moral justifications (Baker & Liu, 2021). Specifically, these preschoolers were more likely to focus on justice and retribution when asked to reason about accidental transgressions. Preschool children's conscience, a set of self-regulatory rules and representations with regard to moral emotions, behavior, and cognition (Kochanska & Askan, 2006), is also longitudinally related to children's aggressive behavior (Kochanska & Knaack, 2003). In general, children's beliefs about the acceptability and utility of aggression predict their overall displays of aggression (Erdley & Asher, 1996, 1998).

Similar to temperament, children's moral reasoning may lessen their need to rely on EF to behave appropriately in conflict situations. Research on the relation between moral reasoning and aggression (e.g., Arsenio et al., 2009; Jambon & Smetana, 2017), as well as on the database in the SIP model (e.g., Burks et al., 1999; Zelli et al., 1999), has shown that children's internalized mental representations can influence behavior in a way that subverts cognitive control. In fact, Burks et al. even referred to the knowledge structures stored in the database as

"distal cognitive control mechanism[s] of aggression" (p. 222), which reinforces the idea that representations of moral knowledge may predispose children towards certain response types and allow for more efficient control.

Unlike temperament, which is thought to maintain relative stability throughout early childhood (Rothbart & Bates, 2006), moral reasoning develops during this time. Yet the process of how children acquire and internalize appropriate moral knowledge is unclear. Thus, a secondary aim of the current study was to explore the factors that contribute to children's moral reasoning about reactive aggression over the preschool year. The constructivist account of moral development posits that children are active participants in the development of their morality through their involvement in early social interactions (Smetana et al., 2018). Young children interpret the messages provided by social situations and construct working mental models of morality. Most acts of harm committed by young children are met with negative reactions from the victim and observers, which provide the transgressor with important information about the act (for review, see Dahl et al., 2018). For example, a 2-year-old child may hit one of her classmates after he takes away her favorite toy. In response, the classmate begins to cry and the teacher tells the child that hitting is not nice because it hurt her classmate. By the end of the incident, the child has received direct feedback about her action from the teacher, as well as indirect feedback from her crying classmate that hitting causes distress. Indeed, the ubiquitous and child-centric nature of peer conflict situations leads to many opportunities for children to engage in this type of active learning. Although previous research has confirmed that children are presented with feedback following transgressions, and considerable development occurs in moral reasoning over the first few years of life (Dahl et al., 2018), there is no research to date that has empirically linked involvement in early conflict situations with later moral reasoning.

In addition, the cognitive mechanisms required for children to be active participants in their moral development have not been empirically examined. What basic cognitive abilities are needed for children to take in and ultimately internalize the relevant information from everyday social situations? To internalize the information presented through the social interactions so that the information can guide later actions, children have to be attentive to both the direct and indirect messages, incorporate those messages into their existing mental models of the social

world, and have the ability to recall those messages at a later time when a similar social situation presents itself. In addition, the type of information that children are presented with may depend on how they act in early social situations, which could also be related to cognitive skills. One cognitive mechanism that is implicated in all of these abilities is EF.

EF assists in the integrative and flexible construction of representations through the use of reflection, the deliberate thought about mental representations, and the iterative reprocessing of information (Zelazo, 2015). When individuals engage in reflection, the contents of thought are elaborated on and combined with previously stored, related information. Done repeatedly throughout early social situations, this process enables the internalization of moral rules that subsequently act as a guide for future social interaction. Thus, EF facilitates the development of moral reasoning by providing children with the mechanism by which they extract moral information from their environments and construct complex, mature moral representations that can be used to inform and guide future behavior.

There is some evidence to suggest that EF is important for moral development. In their research on conscience development, Kochanska and colleagues have found that early cognitive control is an important factor in moral development and the internalization process (Kochanska & Knaack, 2003; Kochanska et al., 1997; Kochanska et al., 1996). For example, Kochanska and Knaack found that children's abilities to inhibit dominant responses (e.g., delay reaching for a desired snack, slowing down motor movement) at 22, 33, and 45 months of age predicted their conscience development at 56 months. Other groups of researchers have found a similar positive relation between EF and moral development (e.g., Baker et al., 2021; Stifter et al., 2009; Vera-Estay et al., 2016).

Smetana et al. (2012) found mixed associations between control abilities at 2.5 to 4.3 years of age and moral reasoning one year later. At initial testing, children with higher control abilities had a better understanding that moral violations were wrong across all situations. However, these same children showed slower growth over the year on other aspects of moral reasoning, including understanding that moral rules are inalterable and that they exist independent of authority. In this sense, children with lower EF appeared to have an advantage in moral

development and the authors hypothesized that this may be because children with lower EF are more likely to be involved in moral transgressions and consequently have more of an opportunity to learn moral rules. However, the moral situations evaluated by children were prototypical, unprovoked transgressions. It is possible that different results would be observed if children are asked to reason about retaliatory events, which may be more complicated and more demanding of EF (Richardson et al., 2012). Indeed, Baker et al. (2021) found that EF related to preschooler's judgments of multifaceted moral transgressions.

The differences in the longitudinal results of Smetana et al. (2012) and Kochanska and colleagues (e.g., Kochanska & Knaack, 2003; Kochanska et al., 1997; Kochanska et al., 1996) may be due to methodological differences in the measurement of moral development. Despite these differences, both groups of researchers discuss the importance of early EF abilities within the context of social situations for the development and internalization of moral rules. Kochanska and colleagues assert that early control abilities in the context of the parent relationship support the development of the conscience because parent relationships offer many opportunities to practice and internalize the directives of others. These practice opportunities within the parenting context provide a foundation for later control abilities in moral contexts and the internalization of moral rules. Smetana et al. (2012) focused more on the frequency that children encounter morally relevant social situations, but also reasoned that even if high EF children are not active participants in moral events as often, they may be better at attending to direct feedback from authorities and taking note of indirect moral messages in their environment.

However, it is likely that both the process outlined by Smetana et al. (2012) and the process outlined by Kochanska and colleagues (e.g., Kochanska & Knaack, 2003; Kochanska et al., 1997; Kochanska et al., 1996) contribute to the development of moral reasoning. The frequency of involvement in moral events matters, but so does the solution of these events. Stable, trait-like EF abilities help children practice moral solutions when faced with peer conflict situations, as well as attend to, interpret, and ultimately internalize the moral information provided from these situations. In this regard, the proposed dissertation study extends the work previously done by Kochanska and colleagues and Smetana et al., by explicitly linking children's EF with moral development through their involvement and resolution of peer conflict in the classroom.

There are very few studies that have examined the relation between EF, moral or social knowledge, and social behavior. Arsenio et al. (2009) found that adolescents that showed high levels of reactive aggression also showed maladaptive social cognitions regarding aggression, but not necessarily deficits in moral reasoning. However, the concurrent link between reactive aggression and social cognitions was mediated by attention issues. The authors suggested that reactive aggressive adolescents likely have difficulty attending to messages in their social environment which contributes to and reinforces maladaptive social cognitions. Despite the lack of a relation between moral reasoning and reactive aggression in their study, their results support the notion that EF abilities influence the type of social information that is internalized and that social cognitions relate to behavior.

In a sample of preschool children, Denham and Bassett (2018) found that EF's concurrent association with emotion knowledge partially mediated the concurrent relation between EF and children's competent responses to hypothetical peer conflict situations. Denham and Bassett suggested that early regulatory abilities may help children attend to emotion information in their environment, which then leads to greater SC. Their line of reasoning, along with that of Arsenio et al. (2009), is remarkably similar to the line of reasoning used in the current study. However, no previous studies have examined how early control abilities interact with children's social behavior to contribute to the development of moral reasoning during the preschool period.

The Current Study

The primary goal of the current study was to examine how EF contributes to the development of SC during the preschool years and how temperament and moral reasoning affect this relation. The secondary goal was to examine how EF and children's classroom behavior contribute to the development of moral reasoning. The examination of how SC, EF, temperament, and moral reasoning relate to one another during the preschool period provides a novel contribution to the literature, as no previous theories or studies have integrated all four constructs.

In the current study, preschool children were tested at the beginning (i.e., September-October) and end (i.e., July-September) of the preschool year. These time points captured children amid

the normative decrease in aggression observed between 3.5 and 5 years of age (e.g., Alink et al., 2006), as well as the rapid period of EF development (e.g., Zelazo et al., 2013). In addition, peer play is a dominant activity in preschool settings and is a critical avenue for the development of social competence (e.g., Cohen & Mendez, 2009; Coolahan et al., 2000). Pre-kindergarten curriculums tend to focus on social-emotional competence and the establishment of the social skills needed to be successful in the kindergarten classroom environment. This is evident by the inclusion of social and emotional development as a foundational element for state-funded programs (e.g., Center on the Social and Emotional Foundations for Early Learning) and the many different curriculum programs that have been designed to specifically address these skills in the preschool classroom (e.g., Fun FRIENDS, Pahl & Barrett, 2010; I Can Solve Problems, Shure & Spivack, 1982; Incredible Years, Webster-Stratton et al., 2008; PATHS, Domitrovich et al., 2007). The presence of these messages in the classroom may help support the development of moral understanding during this time and may also structure activities so that children have more opportunities to develop the skills necessary to navigate peer conflict.

At both time points, participants were given a measure of moral reasoning, a measure of SC in peer conflict, and an EF task battery. Parents completed a temperament questionnaire at the time of consent and teachers completed a questionnaire regarding children's classroom behavior at the end of the school year (i.e., May-July). Based on the importance of having a SC concept outlined by the SIP model (Crick & Dodge, 1994) and Tremblay (2010), it is expected that Time 1 (T1) EF will have an independent, positive association with Time 2 (T2) SC, above and beyond the association between T2 EF and T2 SC. In addition, the longitudinal association between EF and SC will be asymmetrical, meaning that T1 EF will predict T2 SC but T1 SC will not predict T2 EF. Together, this pattern of results would be indicative of an emergence account of the relation between EF and SC because it would illustrate that a certain level of EF competency is required for the development of a SC construct. Further, there will be an interaction between EF and children's temperament or moral reasoning (or both), such that EF will only be associated with SC for children with high surgency, high negative affect, and/or low moral reasoning. This prediction was tested both concurrently and longitudinally, partially because of the low number of participants in the longitudinal sample due to disruptions caused by COVID-19.

Finally, an interaction between T1 EF and children's classroom behavior will predict T2 moral reasoning. Specifically, EF will predict children's moral reasoning, but only if they have had experience with peer conflict and other challenging social situations in the classroom. This prediction is supported by the constructivist account of moral development (e.g., Dahl et al., 2018; Smetana et al., 2018) and would provide converging support for the findings that children's engagement in conflict (Smetana et al. 2012) and their EF (Kochanska & Knaack, 2003) both contribute to children's moral development.

CHAPTER II: METHOD

Participants

Data was collected from 86 pre-kindergarten children at T1 from 12 childcare locations in Greensboro, North Carolina (n = 46 girls; M_{age} = 53 months, SD_{age} = 4 months). The sample size was based on previous longitudinal studies with similar constructs (e.g., Kochanska & Knaack, 2003). Parent-reported demographics indicated that the sample was 54% White, 21% Black, 8% Asian, 8% Multi-Racial, and 9% did not report a race; 6% of the sample additionally identified as Hispanic. Of the 64 participants whose parents reported annual household income, 22% reported earning less than $40,000 a year, 34% reported earning $40,000-$90,000 a year, and 44% reported earning more than $90,000 a year. At T1, children were tested in their childcare centers in Greensboro, NC and surrounding areas. Participants received stickers for their completion of the study tasks at T1.

I originally proposed to test the same children in their childcare centers for T2 during April and May of 2020. One participant was tested in March 2020 in person in anticipation of COVID-related closures, but shutdown occurred before any other participants could have been tested in person. Out of the original 86 participants, I was able to recruit successfully a total of 33 participants for T2 testing (17 girls, M_{age} = 64 months, SD_{age} = 4 months; n = 2 participants received a $10 gift card following participation). Parent-reported demographics indicated that this subset of the sample was 67% White, 15% Black, 9% Asian, 6% Multi-Racial, and 3% did not report a race; 3% of the sample additionally identified as Hispanic. Of the 25 participants whose parents reported annual household income, 8% reported earning under $40,000 a year, 28% reported earning $40,000 - $90,000 a year, and 64% reported earning more than $90,000 a year. A series of chi-square analyses revealed that the T2 participants did not significantly differ on any demographic indices from the participants who only completed T1 (all ps > .10).

The participants' teachers were asked to fill out a questionnaire on the participants' classroom behavior at T2. Sixteen teachers for 69 participants completed the questionnaires and were given $10 in compensation for every questionnaire they complete.

Procedure

Parents completed the temperament questionnaire at the time of consent before the beginning of testing. T1 testing took place between September 2019 and October 2019 and was conducted in person in each participant's childcare center. Participants were given in a fixed order: a peer conflict task, a moral reasoning interview, and a battery of EF tasks (order of EF tasks was randomized). The fixed order procedure was necessary to ensure that participants were not primed to consider the moral implications of their responses on the peer conflict task. In addition, completion of EF task batteries prior to peer conflict tasks may make children more likely to choose aggressive responses to conflict (Caporaso et al., 2019; Caporaso & Marcovitch, 2021). After the completion of each task, children received a sticker as a way of motivating them to complete all the tasks for the study.

T2 testing took place between June 2020 and September 2020. Participants (except for the one mentioned earlier) were tested in their individual homes via Zoom, a virtual meeting platform. Participants were shown each of the tasks on Qualtrics Survey Software using the screen share feature in Zoom. When applicable and possible, participants used the remote-control feature of Zoom to select their answers to the tasks on their devices. Some participants only provided verbal answers because they used devices that were not compatible with the remote-control capabilities in Zoom or were unfamiliar with using a mouse to navigate the screen (the number of participants that provided verbal responses is listed for each specific task). Testing was divided into two separate sessions to accommodate the diminished attention spans of the participants when tested using online methods. This decision was based on a brief piloting period with 4- to 5-year-old children and feedback from their parents. During session one, participants completed the peer conflict task and then the moral reasoning interview in a fixed order. Participants completed the EF task battery during session two. Similar to T1, the individual EF

tasks were presented in a randomized order. Teachers completed the online questionnaire between May 2020 and July 2020.

Measurements

The Virtual School Game (T1, In Person)

The virtual school game (VSG) is an adapted version of Denham et al.'s (2013) challenging situations task and was used to assess children's social problem solving abilities in peer conflict situations. The VSG includes six conflict situations from the challenging situations task (version A) and three added benign situations so participants do not constantly face negative situations during their virtual day at school (see Appendix A for a list and description of the situations).

The VSG was presented on a touchscreen tablet (iPad Air) via Qualtrics Survey Software and was comprised of pictures of the social situations that were created using ToonDoo (Jambav, 2012), a web-based comic program. All situations were presented from the first-person perspective, such that the only character shown on the screen is the transgressor in each situation, and were gender-matched to the participants. Participants were shown pictorial representations of six forced-choice response options, created in Adobe Illustrator, following the presentation of each situation- three aggressive (i.e., physical aggression, verbal aggression, relational aggression) and three competent (i.e., prosocial, avoidant, and tell the teacher)- in a 3x2 grid (see Figure 1).

Before the game began, participants were trained on the flow of the task, how to respond using the touchscreen feature, and were introduced to each of the response options. Participants were first shown a single shape on the first screen. On the following screen, participants were shown an array of six shapes and were asked to select the shape that matches the one on the previous screen by touching the shape on the screen (e.g., "Which shape is the circle? Use your finger to touch the shape that is the circle.").

After the practice shape trials, participants were shown an example situation from the original Challenging Situations Task. In this situation, participants were told that John was waiting for his turn for the swing in line but Aaron came up and pushed John out of line and took his place. Participants were told that they can choose what they want John to do next. Participants were then shown the 3x2 grid of each of the response choices. The experimenter pointed to and labeled each response choice for the participant (e.g., "John could choose to hit Aaron"). Then, the participants were asked to select each response choice on their own (e.g., "What picture would you click if you wanted John to hit Aaron?"). Participants were corrected if they chose the wrong response choice and were asked to select the response choice a second time. All participants chose the correct response choices before moving on to the next phase of the task. The inclusion of this training session ensured that participants were familiar with the response choices and were able to map the response choices with the pictorial representations.

Following the training phase, children were told that they are playing a school game with other real kids their age. They were told that they will see these kids in the game and that they can choose what to do with them. Participants were presented with the nine situations, one at a time, in a randomized order. Participants saw a pictorial depiction of a situation while the experimenter read a brief description of the situation (e.g., "During centers, Stacy knocked over your block tower"). The experimenter then asked the participants "what do you want to do next" and proceeded to a second screen that presented participants with six response options, presented in randomized order. The experimenter pointed and labeled each response option and again asked participants "what do you want to do next" to prompt participants to select their response.

Participants received a score for the total number of each response option they chose (i.e., a score for hit, yell, relational, avoidant, tell the teacher, and prosocial) across the six conflict situations, with a possible range of 0 to 6 for each response type. A competent response composite score was computed by adding up the total number of competent responses they endorsed, collapsed across all three competent response options (i.e., avoidant, tell the teacher, and prosocial). The competent response composite score is the score used in all primary analyses with the VSG.

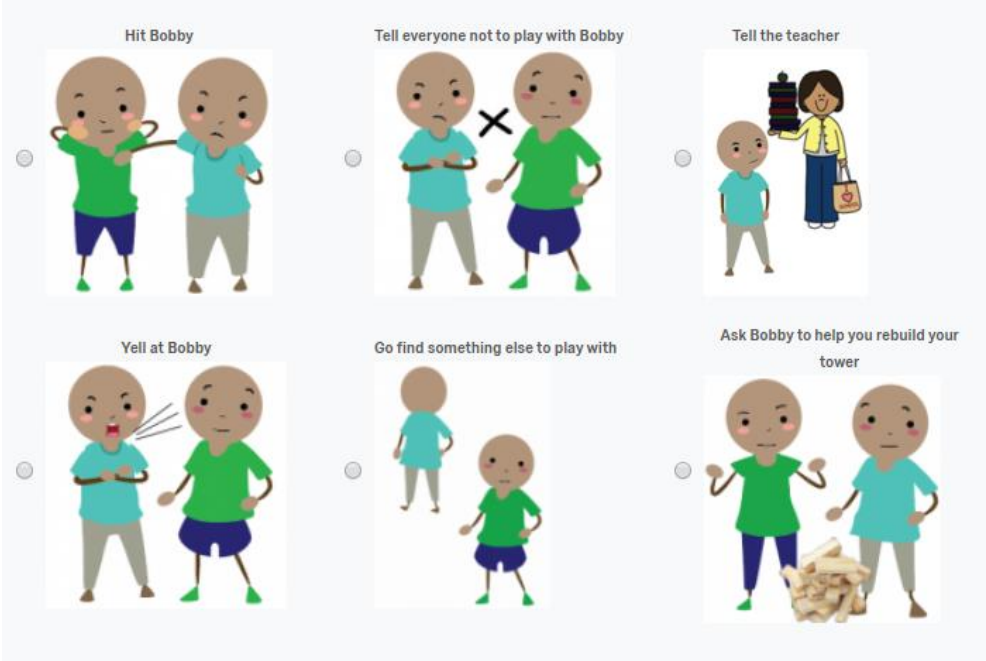

Figure 1. Example social situation on the VSG with response options.

Note. Accompanying script: (top panel) During centers, Bobby knocked over the block tower you had just built (bottom panel) What do you want to do next? Do you want to [label each situation left to right, top to bottom]? What do you want to do [next]/[after Bobby knocked over your block tower]? Experimenters said "next" during in person testing but provided a brief summary of the original provocation during the online version of the task.

The Virtual School Game (T2, Online)

A few adaptations were made to the VSG to accommodate online testing, particularly for participants who provided a verbal response (n = 14). The shape practice trials that were used to orient participants to selecting answers on the screen were skipped because participants who chose to give a verbal response did not need to practice selecting a response on the screen. In addition, the practice trial using the original Challenging Situations Task situation was adapted for these participants. Participants were shown the situation and were introduced to each response option, but instead of being asked to touch each of the response options, they were asked "what would you like John to do next?" and were prompted to provide one verbal answer. To ensure that the participants who continued to provide nonverbal answers intentionally chose their response options, the experimenter repeated their choice in the form of a question after each response selection (e.g., "Tell the teacher?"). The experimenter gave the participant a time to provide a confirmatory response, either verbal (e.g., "yes") or physical (e.g., a head nod).

Finally, the VSG script for all participants was changed slightly. Following the presentation of each situation and the labeling of each response choice, participants were asked what they want to do and were provided with a brief description of the situation (e.g., "What do you want to do after Stacy knocked over your block tower?"). This was done to ensure that participants remembered the original provocation situation after being told all of their response options.

Social Events Interview (T1, In Person)

The social events interview (adapted from Smetana & Ball, 2018) was used to assess moral reasoning. The social events interview was presented on a touchscreen tablet (iPad Air) via Qualtrics Survey Software. Stimuli were comprised of three illustrated vignettes using Dreamstime images in which a target character responds to peer conflict with reactive aggression. The vignettes were gender-matched to the participants and all images were edited in Adobe Illustrator so that every character had a racially ambiguous skin tone and a neutral expression. The three peer conflict scenarios were all acts of unprovoked resource violation, two showed product destruction (i.e., rip a drawing, knock over a block construction) and the other

showed product inequity (i.e., take away a toy). These violations are particularly salient during the preschool years and are the source of the majority of the peer conflicts observed in preschool classrooms (Chen et al., 2001). For each unprovoked act, the target character responded in one of three ways: physical aggression, verbal aggression, or relational aggression. The pairing of each unprovoked act and response option were counterbalanced across the three vignettes (see Table 1 for example). Participants selected their answers to each question of the interview from a 5-point scale using the touchscreen feature. The scale was accompanied by visual representations of each answer that used color, size, and thumbs up/thumbs down as visual cues (see Figure 2).

Table 1. Example pairing of unprovoked acts and retaliatory responses

Unprovoked Act	Retaliatory Response
Rip a drawing	Tell everyone not to play with the transgressor (relational aggression)
Knock over a block construction	Hit the transgressor (physical aggression)
Take away a toy	Yell at the transgressor and say "you better not do that again or else" (verbal aggression)

The social events interview began with a brief scale training procedure, during which participants were shown the different response options to the questions and practiced selecting response options by touching the desired option on the screen (e.g., "Which button would you press for "very good?"). Vignettes were then presented to the participants one at a time, in a randomized order, in a series of pictures (see Figure 2). The first few pictures depicted the unprovoked resource violation act, followed by a picture that depicted the reactive aggression response. Along with the presentation of each picture, the experimenter provided a verbal description of the vignette (see Appendix B for full script). Following each vignette, participants were asked four evaluative questions about the target character's response. Participants were asked about the acceptability of the act, if the act deserves punishment, and two criterion questions regarding rule and authority independence (see Appendix B for the full list of questions and answer scales).

The social events interview has the potential to generate multiple scores; the ratings from each question type can be averaged across stories, ratings for each story can be averaged across question type, etc. For the primary analyses, the average ratings for the two criterion questions (i.e., rule and authority independence) across all three stories were used as the score. This score was chosen because social domain theory suggests that these criterion judgments are the main indicator of the maturity of children's moral concepts (Smetana et al., 2014).

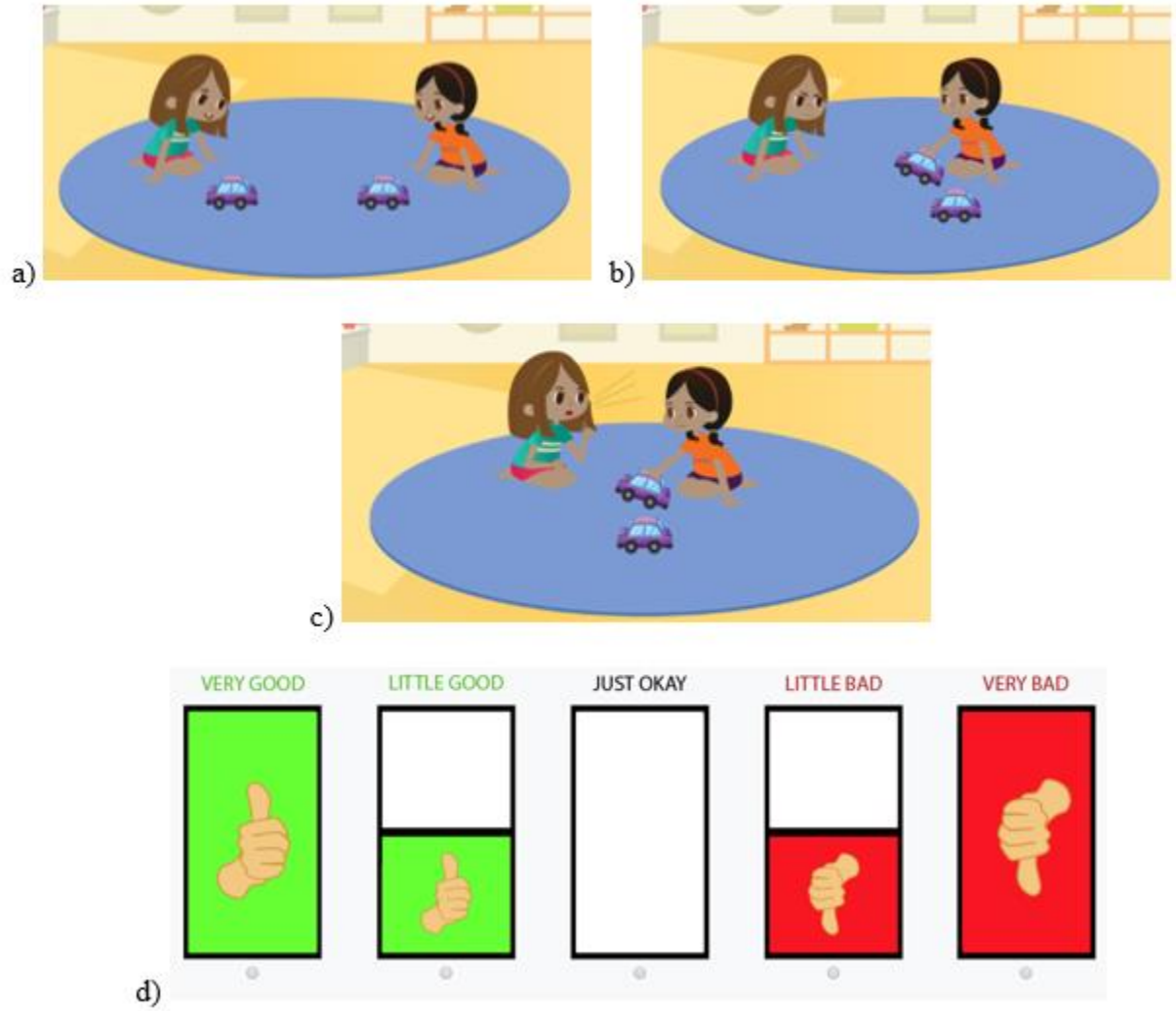

Figure 2. Example vignette, response options, and scale for the social events interview

Note. Accompanying script: a) One day, Jenna and Tina were playing with toy cars. b) Jenna took Tina's car, leaving Tina with no cars to play with. c) When Tina saw that Jenna took her car,

Tina yelled at Jenna and said "you better not do that again or else" (verbal aggression). d) How good or bad do you think it was for Tina to yell at Jenna after Jenna took Tina's car?

Social Events Interview (T2, Online)

Similar to the online version of the VSG, the social events interview was adapted to accommodate verbal responses (n = 14). During scale training, the experimenter labeled each of the scale responses (e.g., "very good", "a little bad") and instructed participants to provide a verbal answer to each question using one of those responses. Also like the online VSG, the experimenter repeated the participants' nonverbal answers in the form of a question (e.g., "a little good?"). No additional changes were made to the social events interview.

EF Task Battery

Dimensional Change Card Sort (T1, In Person)

The NIH Toolbox version of the Dimensional Change Card Sort (DCCS; Zelazo et al., 2013) was used to measure cognitive flexibility and was presented on a touchscreen tablet (iPad Air). The NIH Toolbox version of the DCCS consists of four phases: practice, pre-switch, post-switch, and mixed trials. In all phases, participants were first presented with a fixation screen with a star in the middle to direct attention. Then, the program audibly instructed participants to sort cards that match the target stimuli on one of two dimensions (e.g., color and shape) by touching the picture of the target card they wish to select.

During the practice phase of the task, participants were instructed to sort four cards that match the target stimuli first by the shape dimension, followed by the color dimension. The experimenter demonstrated how to select the answers and performance feedback was provided during this phase. If the participants correctly sorted three out of four cards by shape and color, the participant proceeded to the test trials. The test trials were structured the same way as the practice trials, but with different stimuli and without providing feedback or demonstrations to the participants. Participants were told the sorting rule (e.g., "sort by color") at the beginning of each

phase, but the rule was not repeated before every trial. The test trials began with the pre-switch phase, during which participants were instructed to sort five cards by color. The test trials then proceeded to the post-switch phase, during which participants were instructed to sort five cards by shape. If participants correctly sorted four out of five cards during the post-switch phase, they moved onto the mixed trials phase. During the mixed trials phase, shape and color trials were intermixed; shape trials were presented 75% of the time and color trials were presented 25% of the time. The program audibly told the participants which dimension to sort by before each of the 30 trials.

Scoring of the DCCS was based on both accuracy and reaction time and ranged from 0 to 10, with 5 points provided for accuracy and 5 points provided for reaction time. If participants acquired an accuracy score less than or equal to 80%, then their reaction time was not considered as part of their overall score.

DCCS (T2, Online)

The NIH version of the DCCS was not compatible with online testing. Instead, the DCCS-Borders (Zelazo, 2006) was adapted for online administration via Zoom's screen share feature and Qualtrics Survey Software. Participants were first shown two bivalent target images (i.e., a blue boat and a red bunny). Similar to the NIH version of the DCCS, they were told to select the target image (i.e., a red boat and a blue bunny) that matched the trial item on one of two dimensions (i.e., shape or color). The target images were numbered (i.e., "1", "2") for participants who needed to provide a verbal response (n = 15). Participants who provided a verbal response were instructed to tell the experimenter the target image number that matched the trial image (e.g., "If the picture in the center is red, you will say "1" because the red picture is #1").

DCCS-Borders is structured in a way that is similar to the NIH version of the DCCS and is organized into practice, pre-switch, post-switch, and borders phases. The task began with a brief introduction of the two target cards and then the pre-switch sorting rule was introduced (i.e., shape or color). The order of the sorting rules was counterbalanced across participants.

Participants completed two practice trials of the pre-switch sorting rule and then continued onto the six pre-switch trials. The sorting rule was repeated before every pre-switch trial. Following the pre-switch phase, the sorting rule changed (e.g., switched from color to shape) and participants completed six post-switch trials. The sorting rule was not repeated before every post-switch trial. If participants answered 5 out of 6 of the post-switch trials correctly, they proceeded to the borders phase. In these trials, participants were told that if the trial image had a black border around it, they should sort by color, but if the image did not have a black border around it, they should sort by shape. This rule was repeated before every trial. Participants were also reminded of the pre- and post-switch rules at the beginning of the phase (e.g., "Remember, if it's red, touch this picture/say "1" and if it's blue, touch this picture/say "2"). Participants were given two practice trials (one with a border and one without) before beginning the 12 test trials. The borders test trials were comprised of an equal number of each card type (e.g., 3 bunnies with the border, 3 bunnies without the border).

Participants were given a score based on the total number of correct trials across all three test phases (pre-switch, post-switch, borders), with a possible range of 0-24.

Happy/Sad Stroop (T1 and T2, In Person and Online)

The Happy/Sad Stroop task (H/S Stroop; Lagatutta et al., 2011) was used to measure response inhibition and was presented on an iPad Air (in person) or via screen share on Zoom (online) through Qualtrics Survey Software. Participants were either shown a yellow smiley face ("happy face") or a yellow frown face ("sad face") on each trial. Participants first practiced identifying the correct emotion of each face. Then, participants were told that they were going to play an opposite game. They were instructed to say "happy" when they saw a sad face and "sad" when they saw a happy face. Participants completed two guided trials to test their comprehension of the rules (e.g., "When you see this [sad] face, what will you say?") and then completed four practice trials. Participants received feedback following every practice trial. If participants got at least one practice trial wrong, they completed the four practice trials again before moving on. If participants failed the practice trials more than two times, the task ended (n = 11 failed the

practice trials at T1, *n* = 1 failed the practice trials at T2). If participants passed the practice trials, they moved on to the 20 test trials. The test stimuli were presented at random through Qualtrics. Participants received either a "correct" or "incorrect" on each trial based on the first sound they said (e.g., if a participant said "saaaaa happy" when shown a sad face, the answer was counted as incorrect). Scores on the Happy/Sad Stroop reflect the total number of correct test trials (e.g., said "sad" when shown a "happy" face; possible range: 0-20).

Missing Scan Task (T1, In Person)

An adapted version of the Missing Scan Task (MST; Roman et al., 2014) was used to measure working memory. The task began with two practice trials. Participants were shown two randomly selected pictures of cartoon animals, were asked to label the animals, and were instructed to remember them (i.e., "Can you name these animals? Try to remember which animals these are because they are going to go to the zoo and you will not be able to see them anymore."). If participants were unable to name an animal, the picture was removed from the stimuli set and replaced with another animal. If participants said the wrong name of the animal (e.g., "bird" instead of "turkey"), the picture was kept in the stimuli set because it was assumed that the participant will continue to label the picture in a consistent way for the remainder of the task. After the participants labeled the animals, the experimenter waited 10 seconds and then placed both animals in the "zoo" (a box with a laminated picture of a zoo entrance). After 3-4 seconds, the experiment pulled one animal at random out of the zoo, leaving one animal behind, and asked "what animal is still in the zoo?". If participants named the animal that was directly in front of them instead of the animal still in the box, the experimenter asked them to name the animal that was there before but is now missing (e.g., "well the cat is right in front of you, but what animal also went to the zoo and is now missing?"). Feedback was provided after both practice trials.

Test trials begin after participants correctly answered the two practice trials and followed the same procedure as the practice trials. If participants failed to understand the task after two practice trials, the task was discontinued (*n* = 4 at T1; *n* = 0 at T2). Each round of the test phase

consisted of two trials of the same span size, beginning with a span of four animals and increasing by one for each round (with a maximum span of 10). Even though the span size increased with each round, the experimenter continued to leave one animal behind in the zoo. The animal cards were shuffled after every trial before a new set of cards was selected for the next trial. Feedback was provided after every trial, regardless of performance. Participants needed to answer at least one of the trials correctly per round before moving on to the next span size. The task ended after the participants failed both trials of a span size.

Participants were given a score based on the highest span they answered correctly (i.e., got at least one trial correct), with 0= failed practice trials, 1 = passed practice trials, 2 = passed 4-span, through 7 = passed 10-span.

Missing Scan Task (T2, Online)

The MST was adapted for online administration via Zoom's screen share feature and Qualtrics Survey Software. Digital images of each of the animals that all participants were able to name during T1 were used, as well as a digital image of the zoo entrance. The online version of the MST presented the animal images in a fixed randomized order. That is, the animal images for each trial were randomly selected and the same order was shown to all participants. Similarly, the "missing" animal for each trial was also randomly selected but the same animal was missing for each participant.

The MST task procedure was the same as the in person version; participants completed two practice trials followed by test trials, beginning with the 4-span and ending with the 10-span. However, the online version of the MST had one trial per span size rather than two. This change was made to shorten testing time, an important accommodation for online testing with young children. Participants continued to provide verbal answers on this task. The scoring system also remained the same as the in person version; participants received a score based on the highest span size they answered correctly.

Forward and Backward Digit Span (T2 and Online Only)

The forward and backward digit span task (FDS, BDS; Weschler, 2003) were added at T2. The FDS is considered a measure of short-term memory, while the BDS is considered a measure of working memory (Alloway et al., 2006). The BDS was added as an additional measure of working memory after some concerns with the MST following T1. The task was presented verbally to participants via Zoom. The task started with the forward span trials, followed by the backward span trials. Participants were told to repeat the numbers said by the experimenter verbatim ("say exactly what I say") and were given two 2-span practice trials with feedback. Then, participants were given two trials of each span size, beginning with 2-span and ending with 9-span. The forward span trials ended once participants got both trials of the same span size wrong, with no participant making it past the 7-span trials. No feedback was provided after the test trials.

The experimenter then told the participants that they are going to play "a more difficult game" and proceeded to instruct the participants to say the numbers said by the experimenter backward. The experimenter provided the participants with an example answer and then gave the participants two backward trials. Participants were corrected if they got the practice trials wrong and given an additional opportunity to answer the practice trials correctly. The task ended if participants failed to answer the practice trials after two attempts. Before test trials began, participants were reminded that they need to repeat the numbers backward. Test trials proceeded in the same way as the forward span, with two trials of the same span size each round (except for the 2-span, which consisted of 4 backward trials), beginning with the 2-span and ending with the 9-span (but note that no participant made it past the 4-span).

Participants received 0.5 points for each correct trial (i.e., all numbers were said in the correct order). A forward-digit and backward-digit score were given to each participant to reflect the total number of correct trials in each phase of the task (possible range: 0 – 8 for FDS and 0 – 9 for BDS).

CHILD BEHAVIOR QUESTIONNAIRE- VERY SHORT FORM (T1)

The Child Behavior Questionnaire- Very Short Form (CBQ-VSF; Putnam & Rothbart, 2006) is a parent questionnaire that was used as a measure of temperament. The CBQ-VSF consists of 36 questions that measure three broad temperament dimensions: surgency, negative affect, and effortful control. Each item describes reactions to several different situations and parents were asked to respond on a Likert scale ranging from 1 – "extremely untrue of your child" to 7 – "extremely true of your child" and included a "not applicable" response if the parent had not observed the child in a particular situation.

Example items on the surgency subscale include "often rushes into new situations" and "sometimes turns away shyly from new acquaintances" (reverse coded). Example items on the negative affect subscale include "is difficult to soothe when s/he has become upset" and "gets quite frustrated when prevented from doing something s/he wants to do". Example items on the effortful control subscale include "is good at following instructions" and "when drawing or coloring in a book, shows strong concentration" (see Appendix C for a list of all items by scale). The CBQ-VSF generates three scores, one for each temperament dimension, and reflects the average parent rating for each dimension (possible range: 1-7).

CLASSROOM BEHAVIOR (T2)

Teachers were asked to fill out a questionnaire comprised of the Social Competence and Behavior Evaluation Short Form (SCBE-30; LaFreniere & Dumas, 1996) and subscales from the Preschool Taxonomy of Social Situations (PTOPS; Blankemeyer, et al., 2002), both validated measures of SC for preschool children. Both measures were used to assess participants' classroom behavior with their peers. All instructions and items were entered into Qualtrics Survey Software and teachers were emailed a link to the survey for each of the students who participated in the study. Each individual questionnaire was presented in its own block, with the specific instructions for the questionnaire presented before the items. One teacher requested hard copies of the survey, which were dropped off at the childcare center and then retrieved once

completed. The experimenter followed up two times via email if the teacher did not complete the questionnaire within one to two weeks of receiving the surveys.

The SCBE-30 is comprised of three subscales of 10 questions each on the Anxiety/ Withdrawn and the Social Competence subscales and 9 questions on the Anger/ Aggression subscale. Example items include “does not interact in groups” (Anxiety/Withdrawn), “gets into peer conflicts” (Anger/Aggression), and “accepts reasonable compromises” (Social Competence; see Appendix D for full questionnaire). To avoid confusion, the “social competence” subscale will be referred to as “prosocial behaviors” in the current study as the questions primarily capture children’s helping, sharing, and cooperation behaviors. Teachers selected responses on a Likert scale that ranged from 1 – “almost never occurs” to 6 – “almost always occurs”. The SCBE-30 generates three scores, one for each subscale, and reflects the sum of the ratings across the items per subscale (possible range: 10 – 60 for Anxiety/Withdrawn and Prosocial, 9 – 54 for Anger/Aggression).

The two PTOPS subscales used in the current study were Responses to Provocation (10 items) and Reactive Aggression (8 items). Example items include “when peers call this child a bad name” (Responses to Provocation) and “when a peer refuses to play with this child, s/he gets angry and threatens the peer” (Reactive Aggression; see Appendix D for full questionnaire). To answer these questions, teachers selected responses on a Likert scale that ranged from 1 – “this situation is never a problem for the child” to 5 – “this situation is always a problem for the child”. Scores are generated from each subscale and reflect the sum of the teacher ratings (possible range: 10 – 50 for Responses to Provocation, 8 to 45 for Reactive Aggression).

CHAPTER III: RESULTS

Missing Data

Missing data were handled with the multiple imputation procedure in SPSS. The total scores used in analyses, rather than raw item-level data, were imputed for all missing variables. I chose to handle missing data using the multiple imputation technique because it uses linear regression based on covariance matrices among multiple variables to impute the missing data. In doing so, it maintains the relations among the variables entered into the multiple imputation model. Other techniques, such as "hot-decking" (Myers, 2011), are best when there is one variable that is not missing data and is strongly correlated to the variable with missing data. This was not the case for some of the variables in the current study. For example, the teacher questionnaire subscales were correlated with each other, but the subscale scores all had differing correlations with other study variables and some subscales did not significantly correlate with variables outside the questionnaire. All variables were subjected to 10 imputations and the averages of the imputations were used in data analysis.

There was 9% missing data across all measures at both time points. At T1, $n = 3$ participants were missing one EF task each. At T2, $n = 7$ participants were missing at least one EF task. Across T1 and T2, $n = 2$ participants were missing at least one EF task due to technological issues. Another $n = 4$ participants did not complete at least one EF task due to inattention. For T2 only, $n = 2$ participants were missing data from at least one EF task because of parent interference and another $n = 2$ did not complete the second testing session with the EF tasks. Across all participants at both time points, $n = 6$ participants were missing data for the DCCS, $n = 4$ participants were missing data for H/S Stroop, $n = 6$ participants were missing data for the MST, and $n = 3$ participants were missing the F/BDS. There were also a handful of parents and teachers who did not fill out the questionnaires for the participants ($n = 15$ for parent

questionnaires; n = 17 for teacher questionnaires). There were no missing data for the VSG or the social events interview.

Preliminary Analyses

Time 1 Concurrent Analyses

Before running the primary analyses, I first examined whether there were any differences in the study variables based on demographic variables. Notably, the only gender difference at T1 was in parent-rated surgency; parents rated boys (M = 4.77, SD = 0.89) higher in surgency than girls (M = 4.26, SD = 0.90), $t(84) = 2.60$, $p = .01$. All other comparisons by gender were not significant, all ps > .05. This was a bit surprising given the noted gender differences in aggressive responding during this time (e.g., Crick et al., 2006; Ostrov et al., 2014; Poland et al., 2016). A series of one-way ANOVAs revealed that there were no significant differences in any of the primary analysis variables by race or ethnicity, all ps > .05. Bivariate correlations with T1 age in months and household annual income revealed that age in months was only correlated with performance on the DCCS, $r(86) = .28$, $p = .01$. Income was not significantly correlated with any of the primary analysis variables, all ps > .05.

Time 2 Concurrent Analyses

There were no significant gender differences for the T2 study variables, all ps > .05. Similarly, there were no significant differences by race, and neither income nor age in months at T2 were significantly correlated with any of the primary analyses variables, all ps > .05. Finally, I checked to see whether the mode of response (verbal or non-verbal) affected performance on the VSG, social events interview, and the DCCS through a series of independent sample t-tests. There were no performance differences on any of the measures based on the mode of response, all ps > .05.

Descriptive Analyses

The following sections present the means, standard deviations, and observed range for each measure, frequency tables for the VSG, correlations between the measures, and comparisons between T1 and T2. Results were considered significant at $p < .05$. Results with $p < .10$ were considered marginal and are noted throughout.

Both the individual EF tasks and a composite EF score are reported below. A composite score was used for EF because it is thought that EF is a unitary construct during the preschool period (for review, see Garon et al., 2008). Even if the EF components are separable during this time, they still relate to one another and capture the same underlying control mechanisms. This is supported by the observed correlations between the EF tasks at T1 and T2. It can also be argued that the tasks used in the current study require the use of multiple EF components rather than solely the target component for the task. Indeed, the best practice to examine the contributions of individual EF components would be to include multiple measures for the same component to capture the specific variance associated with each individual component. To create the EF composite score, the scores from the individual EF measures were standardized and then the mean of the standardized scores was calculated. At T1, all three EF variables were correlated and were used to make the composite EF score. The T2 EF composite was comprised of the DCCS, MST, and BDS. The FDS was not included in the composite score because it is considered to be a measure of short-term memory and not working memory (Alloway et al., 2006). T2 H/S Stroop was excluded from the composite because it was not correlated with the other EF tasks nor was it predicted by any of the T1 EF measures.

For each of the bivariate correlation analyses, five interaction variables of interest were included based on the a priori hypotheses: EF x surgency, EF x negative affect, criterion judgments x surgency, criterion judgments x negative affect, and EF x criterion judgments. In addition, four interactions with gender were included: gender x EF, gender x criterion judgments, gender x surgency, and gender x negative affect. There were no observed gender differences in competent responses, but because gender differences are ubiquitous in the aggression literature (e.g., Crick

et al., 2006; Ostrov et al., 2014; Poland et al., 2016), these interactions were included. All variables were centered before the computation of the interaction variables.

Time 1 Concurrent Analyses

Table 2 contains the means, standard deviations, and observed range for the T1 variables. The table includes all individual and composite scores for all measures, including the parent questionnaire. Table 3 lists the frequencies of competent responses on the VSG. Notably, 20.9% of the sample chose competent responses for all conflict situations at T1. Differences on all measures were examined between the longitudinal sample subset (n = 33) and the participants who did not complete T2 (n = 52) using a series of independent samples t-tests. There were no significant differences between the two sets of participants at T1, all ps > .10.

Table 4 lists the zero-order bivariate correlations between the T1 variables used in the primary analyses. Of note, competent responses were significantly, and positively, associated with criterion judgments, $r(86) = .282$, $p = .009$, and EF composite scores, $r(86) = .296$, $p = .01$. The three EF measures were all correlated with one another but only the H/S Stroop was significantly, positively, correlated with competent responses, $r(86) = .370$, $p < .001$. The temperament dimensions were not correlated with any of the other variables as main effects, but the criterion x negative affect interaction was significantly correlated with competent responses $r(86) = .227$, $p = .04$. Two significant interactions with gender were also correlated with competent responses: gender x surgency, $r(86) = -.216$, $p = .045$ and gender x negative affect, $r(86) = .230$, $p = .03$.

T1 correlations were also examined with the longitudinal sample subset, the participants that participated in both T1 and T2 (n = 33) as many participants were unable to be tested at T2 due to COVID-19 shutdowns and complications. Three notable differences between the full sample and the longitudinal subsample at T1 emerged. First, there was a significant positive correlation between EF composite scores and criterion judgments $r(33) = .404$, $p = .02$. Second, there was a significant negative correlation between negative affect and competent responses, $r(33) = -.349$,

p = .047. Finally, neither of the gender x temperament interactions were significantly correlated with competent responses, both ps > .10.

Table 2. Descriptive statistics for T1 variables

Variable	Mean	Standard Deviation	Observed Range
VSG			
Physical Aggression	0.71	0.92	0 to 5
Verbal Aggression	0.48	0.72	0 to 3
Relational Aggression	0.61	0.79	0 to 3
Avoidant	0.56	0.90	0 to 6
Tell Teacher	0.89	1.04	0 to 6
Prosocial	2.80	1.67	0 to 6
Total: Competent Responses	4.21	1.44	0 to 6
Social Events Interview			
Acceptability	4.26	0.95	1 to 5
Deserved Punishment	3.04	1.07	1 to 5
Authority Independence	3.68	1.04	1 to 5
Rule Independence	3.57	1.17	1 to 5
Physical Aggression	3.86	1.04	1 to 5
Verbal Aggression	3.84	1.04	1 to 5
Relational Aggression	3.82	1.07	1 to 5
Total: Criterion Judgments	3.62	0.93	1 to 5
EF Tasks			
DCCS	2.61	1.85	0 to 6.42
H/S Stroop	12.24	5.84	0 to 20
MST	2.41	1.27	0 to 5
Total: EF Composite	0	2.13	-5.41 to 4.16
Temperament			
Surgency	4.50	0.92	2.09 to 7.00
Negative Affect	4.13	0.80	2.08 to 5.73
Effortful Control	5.19	0.70	2.60 to 6.25

Table 3. Frequencies of competent responses on the VSG at T1

# Competent Responses	*n*	%
0	1	1.2%
1	4	4.7%
2	5	5.8%
3	14	16.3%
4	22	25.6%
5	22	25.6%
6	18	20.9%

Table 4. Correlations between T1 primary analyses variables.

	1.	2.	3.	4.	5.	6.	7.	8.	9	10	11	12.	13.	14.	15.	16.	17.
1. Comp. Responses	1	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
2. Criterion Judgments	.282**	1	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
3. EF Composite	.296**	.186^{+}	1	--	--	--	--	--	--	--	--	--	--	--	--	--	--
4. DCCS	.115	.125	.730**	1	--	--	--	--	--	--	--	--	--	--	--	--	--
5. H/S Stroop	.370**	.144	.691**	.256*	1	--	--	--	--	--	--	--	--	--	--	--	--
6. MST	.146	.129	.711**	.299**	.217*	1	--	--	--	--	--	--	--	--	--	--	--
7. Surgency	-.178	-.183^{+}	-.130	-.044	-.118	-.116	1	--	--	--	--	--	--	--	--	--	--
8. Negative Affect	-.025	.112	-.058	.009	-.161	.028	-.086	1	--	--	--	--	--	--	--	--	--
9. Effortful Control	.044	.103	.077	.020	.076	.067	-.239*	.139	1	--	--	--	--	--	--	--	--
10. EF x Surgency	.060	.132	.057	.110	-.028	.040	.054	.107	.097	1	--	--	--	--	--	--	--
11. EF x NegAff	.125	.116	.076	-.042	.033	.171	.119	-.155	.028	-.175	1	--	--	--	--	--	--
12. Criterion x Surgency	.065	.174	.125	-.001	.028	.241*	.017	-.136	.067	.096	.087	1	--	--	--	--	--
13. Criterion x NegAff	.227*	-.001	.111	.092	.100	.044	-.153	.098	.042	.114	-.127	-.050	1	--	--	--	--
14. EF x Criterion	-.086	-.173	-.041	.008	-.141	.045	.132	.104	-.023	-.008	.200^{+}	-.041	-.200^{+}	1	--	--	--
15. Gender x EF	.073	-.122	.026	-.005	.134	-.074	-.087	-.139	-.211^{+}	-.224*	.088	-.007	-.118	.097	1	--	--
16. Gender x Criterion	.146	-.259*	-.120	-.045	-.068	-.143	.026	-.041	-.138	.008	-.117	-.305**	.134	.039	.202^{+}	1	--
17. Gender x Surgency	-.216*	.027	-.091	-.013	-.098	-.082	.003	-.081	.181^{+}	.059	.006	.180	-.108^{+}	.047	-.125	-.161	1
18. Gender x NegAff	.230*	-.042	-.139	-.190	.077	-.183	-.078	-.023	.133	.016	.078	-.128	.241*	-.105	-.039	.120	-.088

Note. $^{+}$ $p < .10$, * $p < .05$, ** $p < .01$

Time 2 Concurrent Analyses

Table 5 depicts the means, standard deviations, and observed range for T2 variables. The table includes all individual and composite scores for all measures, including the classroom behavior variables. Table 6 lists the frequencies of each response type on the VSG. Notably, 54.5% of the sample chose all competent responses at T2 and no participant chose less than 3 competent responses. Table 7 lists the correlations between T2 competent responses, T2 criterion judgments, T2 EF composite and individual tasks, and seven interactions: T2 EF x surgency, T2 EF x negative affect, T2 criterion x surgency, T2 criterion x negative affect, T2 EF x T2 criterion, gender x T2 EF, and gender x T2 criterion. Of note, there was a significant positive correlation between competent responses and EF composite scores, $r(33) = .723$, $p < .001$. Interestingly, the H/S Stroop was the only EF task that was not correlated with competent responses, $r(33) = .246$, $p = .17$. H/S Stroop was also not significantly correlated with any of the other EF tasks, all ps $> .05$. Criterion judgments were not significantly correlated with any other T2 variable. The T2 EF x surgency interaction was the only interaction significantly correlated with T2 competent responses, $r(33) = .504$, $p = .003$.

Table 5. Descriptive statistics for T2 variables

Variable	Mean	Standard Deviation	Observed Range
VSG			
Physical Aggression	0.21	0.55	0 to 2
Verbal Aggression	0.24	0.66	0 to 3
Relational Aggression	0.27	0.52	0 to 2
Avoidant	0.36	0.55	0 to 2
Tell Teacher	1.21	1.14	0 to 5
Prosocial	3.70	1.74	0 to 6
Total: Competent Responses	5.27	0.95	0 to 6
Social Events Interview			
Acceptability	4.21	0.78	1.67 to 5
Deserved Punishment	3.04	1.05	1 to 5
Authority Independence	3.40	1.09	1 to 5
Rule Independence	3.21	1.06	1 to 5
Physical Aggression	3.64	1.04	1 to 5
Verbal Aggression	3.56	1.02	1.67 to 5
Relational Aggression	3.64	0.99	1 to 5
Total: Criterion Judgments	3.31	0.99	1.67 to 5
EF Tasks			
DCCS	17.18	4.92	6 to 24
H/S Stroop	15.55	4.31	0 to 20
MST	2.06	0.93	1 to 4
FDS	3.26	0.89	1 to 5
BDS	1.36	1.09	0 to 3
Total: EF Composite	0	0.82	-1.55 to 1.50
Classroom Behavior			
Anxiety/Withdrawn	19.95	6.17	10 to 34
Anger/Aggression	20.14	7.72	9 to 50
Prosocial	42.77	8.10	18 to 60
Response to Provocation	29.07	7.81	11 to 50
Reactive Aggression	18.52	7.66	8 to 40

Table 6. Frequencies of competent responses on the VSG at T2

# Competent Responses	*n*	%
3	2	6.1%
4	5	15.2%
5	8	24.2%
6	18	54.5%

Table 7. Correlations between T2 primary analyses variables.

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.
1. Comp. Responses	1	--	--	--	--	--	--	--	--	--	--	--	--	--
2. Criterion Judgments	.142	1	--	--	--	--	--	--	--	--	--	--	--	--
3. EF Composite	.723**	.071	1	--	--	--	--	--	--	--	--	--	--	--
4. DCCS	.770**	.029	.870**	1	--	--	--	--	--	--	--	--	--	--
5. H/S Stroop	.246	-.040	.377*	.371*	1	--	--	--	--	--	--	--	--	--
6. MST	.406*	.177	.793**	.570**	.248	1	--	--	--	--	--	--	--	--
7. FDS	.231	.124	.423*	.456**	.208	.416*	1	--	--	--	--	--	--	--
8. BDS	.598**	-.032	.792**	.567**	.305^{+}	.376*	.167	1	--	--	--	--	--	--
9. EF x Surgency	.504**	-.191	.092	.286	-.101	-.057	-.134	-.004	1	--	--	--	--	--
10. EF x NegAff	-.158	-.117	-.126	-.158	-.243	-.184	.008	.033	-.236	1	--	--	--	--
11. Criterion x Surgency	-.015	.058	-.196	-.180	-.328^{+}	.026	.021	-.327^{+}	.364*	-.107	1	--	--	--
12. Criterion x NegAff	-.154	-.069	-.112	-.107	-.366*	-.205	-.101	.037	.051	.186	.072	1	--	--
13. EF x Criterion	.103	-.016	.389*	.298^{+}	.112	.411*	.012	.247	-.162	.095	-.444**	-.063	1	--
14. Gender x EF	-.056	-.009	.192	.095	.032	.211	.205	.164	-.207	-.154	.084	-.182	-.348*	1
15. Gender x Criterion	-.033	-.222	-.010	.079	.374*	-.237	.153	.133	-.081	-.190	-.377*	.120	.006	.067

Note. $^{+}$ $p < .10$, * $p < .05$, ** $p < .01$

LONGITUDINAL ANALYSES

Correlations between T2 variables and T1 variables, classroom behavior variables, and interactions between T1 EF and gender, temperament, T1 criterion judgments, and the classroom behavior variables, as well as between the T1 criterion judgments and temperament and gender are presented in Table 8. T2 competent responses were significantly, positively, associated with T1 competent responses, $r(33) = .357$, $p = .04$, T1 criterion judgments, $r(33) = .404$, $p = .02$, and T1 EF composite scores, $r(33) = .590$, $p < .001$. All three T1 EF tasks, and not just H/S Stroop, correlated with T2 competent responses, $r(33)_{DCCS} = .450$, $p = .01$, $r(33)_{H/S\ Stroop} = .378$, $p = .03$, $r(33)_{MST} = .440$, $p = .01$. Surgency was negatively associated with T2 competent responses, $r(33) = -.531$, $p = .001$. Many of the T2 EF variables were correlated with T1 variables, notably T1 competent responses, T1 criterion judgments, and surgency, all *p*s < .05. The only classroom behavior variable that was significantly related to any other variable was prosocial behaviors, which was negatively associated with T2 criterion judgments, $r(33) = -.504$, $p = .003$. It is also important to note that T2 H/S Stroop was not significantly correlated to any T1 variables, including the T1 EF variables, all *p*s > .05.

Several interactions were also significantly correlated with T2 variables. T1 EF x surgency was significantly related to T2 competent responses, $r(33) = .472$, $p = .01$. T1 criterion x surgency was significantly related to T2 criterion judgments, $r(33) = -.365$, $p = .04$. EF x prosocial, $r(33) = -.332$, $p = .06$, and EF x anger/aggression were marginally related to T2 criterion judgments, $r(33) = .320$, $p = .07$. T1 EF x T1 criterion were significantly associated with T2 criterion judgments, $r(33) = .384$, $p = .03$, and T2 EF, $r(33) = -.352$, $p = .04$. Notably, none of the gender interactions with T1 variables were correlated to any of the T2 variables, all *p*s > .10.

Table 8. Longitudinal correlations

	T2 Competent	T2 Criterion	T2 EF	T2 DCCS	T2 H/S Stroop	T2 MST	T2 FDS	T2 BDS
T1 Competent	.357*	.094	.504**	.305^{+}	.062	.456**	.077	.476**
T1 Criterion	.404*	-.079	.456**	.372*	.008	.342*	.192	.404*
T1 EF Composite	.590**	.028	.625**	.558**	.120	.354*	.013	.622**
T1 DCCS	.450**	.133	.353*	.380*	.037	.093	-.087	.393*
T1 H/S Stroop	.378*	.029	.568**	.417*	.145	.409*	.006	.568**
T1 MST	.440*	-.110	.463**	.401*	.076	.257	.116	.370*
Surgency	-.531**	-.037	-.551**	-.548**	-.188	-.483**	-.266	-.322^{+}
Negative Affect	.186	.252	.037	.181	.127	.102	.174	-.193
Effortful Control	.270	.032	.038	.068	-.135	.024	-.025	.003
Anxiety/Withdrawn	-.064	-.020	.038	-.130	-.311^{+}	.283	.044	-.060
Anger/Aggression	-.128	-.142	.181	.010	.266	.290	.116	.143
Prosocial	.110	-.504**	-.141	-.044	.014	-.235	.058	-.068
Response to Provocation	.068	-.042	.76	.206	.241	.220^{+}	.226	.141
Reactive Aggression	.010	.025	.098	.005	.139	.152	.091	.082
T1 EF x Surgency	.472**	-.106	.161	.214	.010	.022	-.063	.157
T1 EF x Neg. Affect	-.047	-.135	.005	-.104	-.176	.024	-.035	.092
T1 Criterion x Surgency	.001	-.365*	.082	.100	.015	-.026	.116	.128
T1 Criterion x Neg. Affect	.044	-.216	.045	.077	-.247	-.181	-.163	.214
T1 EF x T1 Criterion	-.106	.384*	-.353*	-.249	-.144	-.352*	-.210	-.265
T1 EF x Anx/With	.068	.026	.167	.222	-.185	.064	-.065	.123
T1 EF x Ang/Agg	.151	.320^{+}	-.127	.046	.142	-.184	-.077	-.173
T1 EF x Prosocial	-.264	-.332^{+}	-.250	-.170	-.144	-.259	-.307^{+}	-.185
T1 EF x Provocation	.062	.157	-.151	-.038	.083	-.174	.117	-.157
T1 EF x Reactive Agg	.104	.117	-.057	-.001	.016	-.125	.061	-.014
Gender x T1 EF	.058	-.239	.241	.138	.049	.271	.180	.182
Gender x T1 Criterion	-.075	-.021	.040	-.015	.061	.073	-.093	.039
Gender x Surgency	.240	.008	.068	.064	.025	-.053	-.058	.156
Gender x Neg. Affect	-.079	-.281	-.187	-.070	.018	-.340*	-.316^{+}	-.049

Note. $^{+}$ $p < .10$, * $p < .05$, ** $p < .01$

A series of paired samples t-tests were conducted on the T2 and T1 variables to observe the overall difference in the measures between the two time points. Participants chose significantly more competent responses at T2 compared to T1, $t(32) = 3.03$, $p = .01$ and had significantly higher scores on the H/S Stroop at T2 compared to T1, $t(32) = 3.67$, $p < .001$. Performance on the T1 MST was marginally better than T2 MST performance, $t(32) = -1.88$, $p = .07$. Perhaps participants found it harder to remember the animals when they were presented on a computer screen than when they were presented as physical cards in front of them. I was unable to analyze the difference between T1 and T2 DCCS because they are two different versions of the DCCS task with substantially different scoring systems, but the T1 and T2 versions of the DCCS were correlated, suggesting that participants maintained rank-order stability despite task differences. Participants' average criterion ratings on the social events interview did not significantly differ between T1 and T2, $t(32) = -1.51$, $p = .14$. Curiously, the descriptive data suggest that participants selected lower criterion ratings at T2 than they did at T1.

Primary Analyses

Multiple linear regressions were used to test the main hypotheses of the study. Variables were only entered into the models if they held theoretical significance and were correlated (significantly or marginally) with the outcome variable either as a main effect or as part of an interaction. The variables of theoretical interest for competent responses as an outcome measure were EF, criterion judgments, temperament, and gender. The variables of theoretical interest for EF as an outcome measure were criterion judgments, competent responses, and temperament. The variables of theoretical interest for criterion judgments as an outcome measure were EF, classroom behavior, and gender. The stepwise entry method was used in models that exceeded 4 predictor variables to account for multicollinearity issues among the various predictor variables. The criteria for model inclusion using stepwise entry was set to $\alpha = .05$. All significant interactions were followed up using simple slopes tests that were conducted and graphed via a macro in excel (downloaded from www.jeremydawson.co.uk/slopes.htm). For each test, values of the independent variable were plotted and tested at ± 1 SD of the moderator (Aiken & West, 1991).

TIME 1 CONCURRENT ANALYSES

Full Sample

T1 competent responses were regressed on gender, T1 composite EF scores, T1 criterion judgments, surgency, negative affect, and three two-way interactions: T1 criterion x negative affect, gender x surgency, and gender x negative affect (see Table 9). Using the stepwise entry method, EF was entered into the model in step 1 and significantly predicted competent responses on the VSG, $\beta = .30$, $t(84) = 2.84$, $p = .01$, $R^2 = .088$. The gender x negative affect interaction was entered in step 2 and significantly predicted competent responses, $\beta = .28$, $t(83) = 2.73$, $p = .01$, $\Delta R^2 = .075$. Simple slopes tests were conducted to follow up the significant gender x negative affect interaction and revealed that negative affect had a positive association with competent responses for girls, $t(83) = 2.68$, $p = .01$, and a negative association with competent responses for boys, $t(83) = -2.76$, $p = .01$ (see Figure 3). Criterion judgments were entered in step 3 and significantly predicted competent responses, $\beta = .24$, $t(82) = 2.41$, $p = .02$, $\Delta R^2 = .055$. EF and the gender x negative affect interactions remained significant predictors of competent responses throughout.

Table 9. T1 competent responses regressed on T1 EF, T1 moral reasoning, temperament, and gender ($n = 86$).

	Step 1			Step 2			Step 3		
Variables	*B*	*SE*	*B*	*B*	*SE*	*B*	*B*	*SE*	*B*
EF Composite	.20	.07	.30**	.23	.07	.33**	.19	.07	.29**
Gender x NegAff				.99	.36	.28**	1.01	.35	.28**
Criterion Judgments							.37	.15	.24*
F		8.07**			8.06***			7.62***	
R^2		.088			.163			.218	
ΔR^2					.075			.055	

Note. **$p < .01$, ***$p < .001$

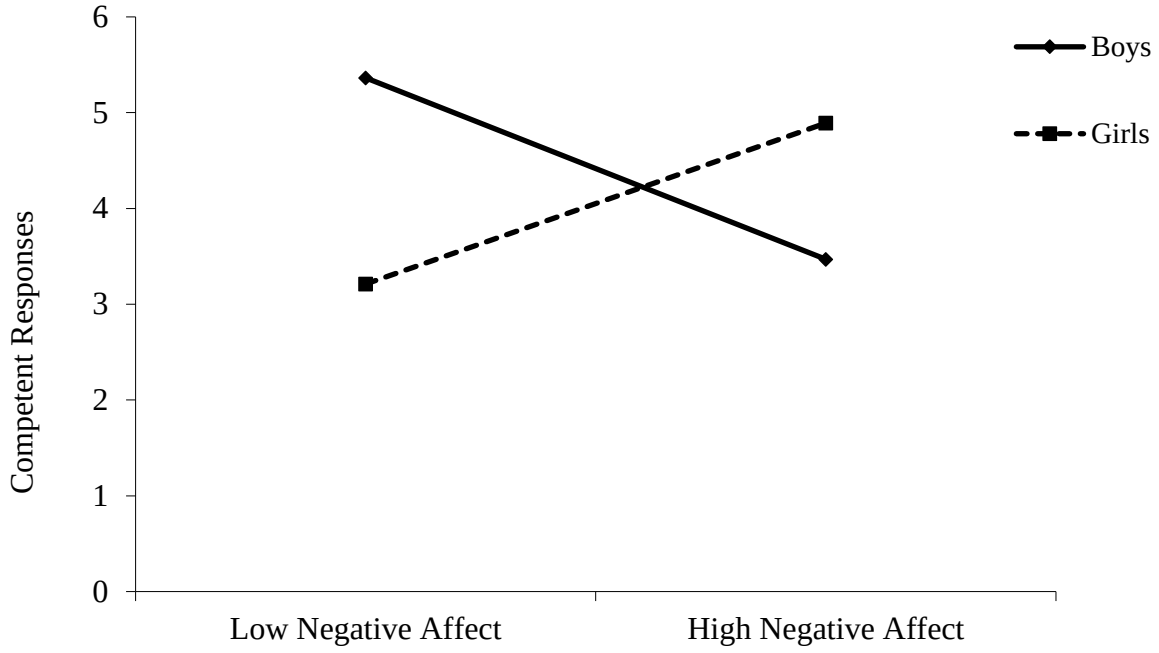

Figure 3. T1 competent responses regressed on the gender x negative affect interaction

Longitudinal Subsample

Similar analyses were run with the longitudinal subsample (n= 33) for the purposes of direct comparison with the T2 concurrent data and to assess whether the subsample is meaningfully different from the full sample at T1. T1 competent responses were regressed on T1 composite EF scores, T1 criterion judgments, negative affect, and T1 criterion x negative affect (see Table 10). EF, β = .47, t(28) = 3.25, p = .003, and negative affect, β = -.34, t(28) = -2.54, p = .02, independently predicted competent responses. The criterion x negative affect interaction marginally predicted competent responses, β = .25, t(28) = 1.83, p = .08. Follow up simple slope tests revealed that criterion ratings significantly predicted competent responses for participants with high negative affect, t(28) = 2.11, p = .05, but the correlation between moral reasoning and competent responding was not significant for participants with low negative affect, t(28) = -0.53, p = .60 (see Figure 4).

These results indicate that the longitudinal subsample differed from the full sample at T1 in a few ways. Although EF was a significant predictor of competent responses in both samples, criterion judgments only predicted competent responses for participants high in negative affect in the subsample. Negative affect also emerged as a significant predictor as a main effect in the subsample but the gender x negative affect interaction was not significant.

Table 10. T1 competent responses regressed on T1 EF, T1 moral reasoning, and negative affect for the T1 longitudinal subsample (n = 33).

Variables	*B*	*SE*	*B*
EF Composite	.34	.11	.47**
Negative Affect	-.61	.24	-.34*
Criterion Judgments	.23	.22	.15
Criterion x Negative Affect	.48	.26	.25^{+}
F	13.64***		
R^2	.532		

Note. + p < .10, * p < .05, ** p < .01, *** p < .001

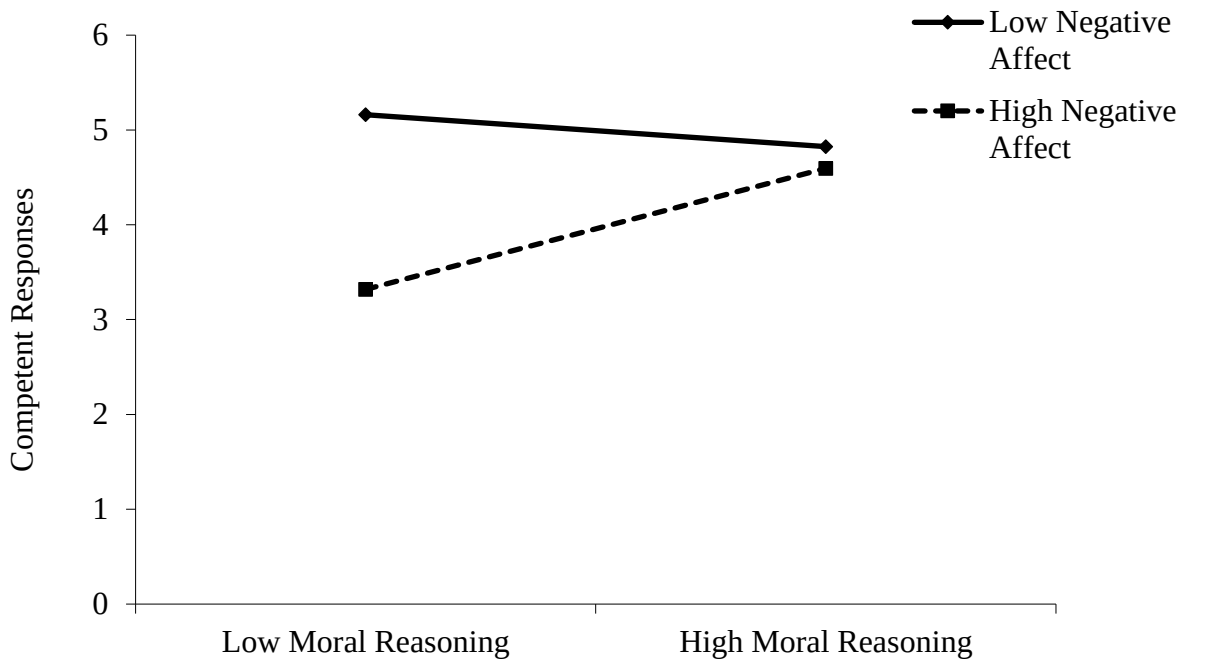

Figure 4. T1 competent responses regressed on the T1 criterion x negative affect interaction for the T1 longitudinal subsample

TIME 2 CONCURRENT ANALYSES

T2 competent responses were regressed on T2 composite EF scores, surgency, and T2 EF x surgency (see Table 11). T2 EF, $\beta = .62$, $t(29) = 5.32$, $p < .001$, and the T2 EF x surgency interaction, $\beta = .43$, $t(29) = 4.29$, $p < .001$, significantly predicted T2 competent responses on the VSG. Follow up simple slopes tests revealed that EF significantly predicted competent responses on the VSG for participants high in surgency, $t(29) = 8.53$, $p < .001$, but the correlation between EF and competent responses is not significant for participants low in surgency, $t(29) = 1.45$, $p = .16$ (see Figure 5). Note that the maximum number of competent responses is 6. The trend observed for high surgency is likely asymptotic rather than linear.

Table 11. T2 competent responses regressed on T2 EF and surgency.

Variables	*B*	*SE*	*B*
EF Composite	.72	.14	.62***
Surgency	-.09	.09	-.11
EF x Surgency	.42	.10	.43***
F	25.25***		
R^2	.723		

Note. + p < .10, * p < .05, ** p < .01, *** p < .001

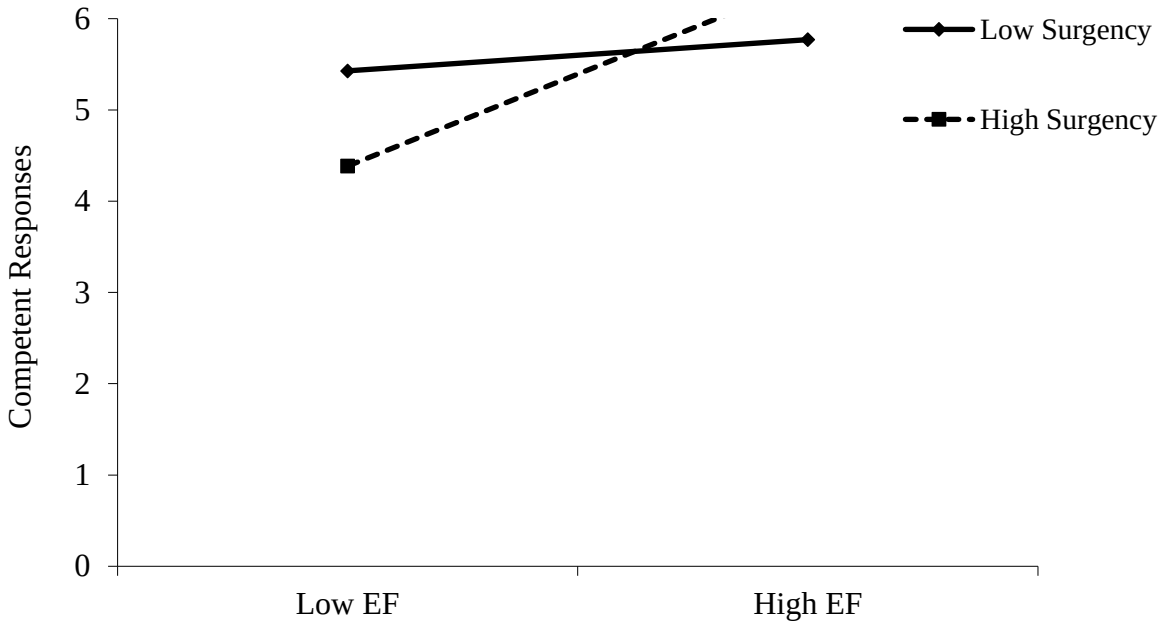

Figure 5. T2 competent responses regressed on the T2 EF x surgency interaction

Interim Summary: Concurrent Results

The T1 concurrent data with the full sample revealed that both EF and moral reasoning independently, positively, predicted competent responses on the VSG. In addition, the gender x negative affect interaction revealed differential associations between displays of negative emotion and SC for each gender; negative affect was negatively associated with SC for boys and positively associated with SC for girls. This interaction was not observed in the longitudinal subsample; the main effect for negative affect negatively predicted SC across genders. Like the full T1 sample, EF predicted SC in the longitudinal subsamples at T1, but criterion judgments were only related to competent responses via their interaction with negative affect. At T2, EF continued to predict competent responses but the T2 EF x surgency interaction indicated that EF was only significantly associated with competent responses for participants high in approach tendencies.

Longitudinal Analyses

The following section outlines the analyses that examine the longitudinal predictors of competent responses on the VSG, as well as EF and criterion judgments, at T2. Hierarchical linear regressions were conducted for each outcome measure. In the first step, the T1 predictor variables were entered along with the T1 outcome variable as a control variable. In the second step, the T2 predictor variables were entered to demonstrate whether the T1 predictor variables have a direct or indirect association with the T2 outcome variable. The stepwise entry method was used.

Competent Responses

T2 competent responses were regressed on T1 competent responses, T1 EF, T1 criterion judgments, surgency, and the T1 EF x surgency interaction in step 1. T2 EF and T2 EF x surgency were then entered in step 2 (see Table 12). T1 EF was entered in step 1 and significantly predicted T2 competent responses, $\beta = .59$, $t(31) = 4.06$, $p < .001$. Surgency was entered in step 2 and significantly, negatively, predicted T2 competent responses, $\beta = -.41$, $t(30)$

= -3.03, p = .005. The T1 EF x surgency interaction was entered in step 3 and significantly predicted T2 competent responses, β = .36, t(29) = 3.12, p = .004. T1 EF and surgency remained significant predictors of T2 competent responses. Follow up simple slopes tests revealed that T1 EF predicted T2 competent responses for participants with high surgency, t(29) = 4.49, p < .001, but the correlation between EF and competent responding was not significant for participants with low surgency, t(29) = 0.31, p = .76 (see Figure 6). T2 EF was entered in step 4, significantly predicted T2 competent responses, β = .42, t(28) = 2.72, p = .01, reduced surgency to marginal significance, and reduced T1 EF to nonsignificance. The T1 EF x surgency interaction remained a significant predictor of T2 competent responses.

Table 12. T2 competent responses regressed on T1 EF, T1 moral reasoning, temperament, and classroom behavior variables.

	Step 1			Step 2			Step 3			Step 4		
Variables	*B*	*SE*	*β*	*B*	*SE*	*B*	*B*	*SE*	*β*	*B*	*SE*	*β*
T1 EF	.27	.07	.59***	.22	.06	.49**	.18	.05	.41**	.09	.06	.19
Surgency				-.32	.11	-.41**	-.32	.09	-.41**	-.19	.10	-.23^{+}
T1 EF x Surgency							.13	.04	.36**	.13	.04	.33**
T2 EF										.48	.18	.42*
F	16.51***			15.04***			16.17***			16.67***		
R^2	.348			.501			.626			.704		
ΔR^2				.153			.125			.078		

Note. + p < .10, * p < .05, ** p < .01, *** p < .001

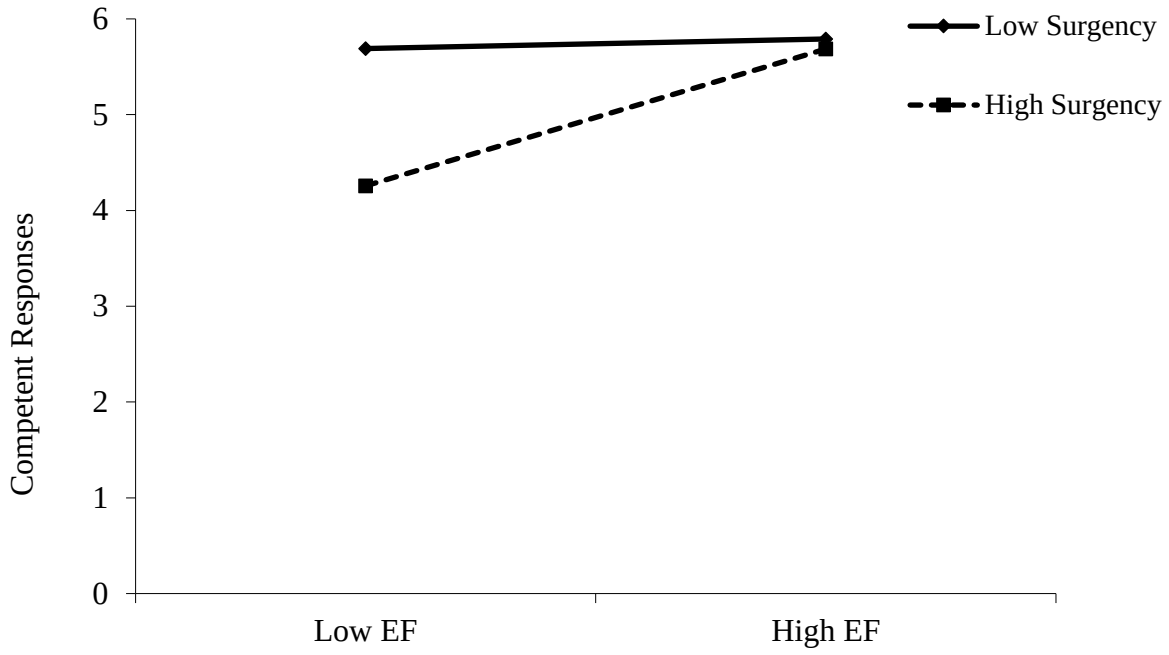

Figure 6. T2 competent responses regressed on the T1 EF x surgency interaction

EF

T2 EF was regressed on T1 EF, T1 competent responses, T1 criterion judgments, surgency, and the T1 EF x T1 criterion interaction in step 1, and T2 competent responses, T2 criterion judgments, and the T2 EF x T2 criterion judgment interaction in step 2 (see Table 13). T1 EF was entered in the step 1 and significantly predicted T2 EF, $\beta = .63$, $t(31) = 4.46$, $p < .001$. Surgency was entered in step 2 and significantly, negatively predicted T2 EF, $\beta = -.42$, $t(30) = -3.31$, $p = .002$. T1 EF remained a significant predictor of T2 EF. Finally, T2 competent responses were entered in step 3 and significantly predicted T2 EF, $\beta = .39$, $t(29) = 2.46$, $p = .02$, and reduced surgency to marginal significance. T1 EF remained a significant predictor of T2 EF.

Table 13. T2 EF regressed on T1 EF, competent responses, and temperament

	Step 1			Step 2			Step 3		
Variables	*B*	*SE*	*B*	*B*	*SE*	*β*	*B*	*SE*	*B*
T1 EF	.25	.06	.63***	.20	.05	.52***	.13	.06	.33*
Surgency				-.29	.09	-.42**	-.18	.09	-.26^{+}
T2 Competent Responses							.34	.14	.39*
F	19.88***			18.58***			16.50***		
R^2	.391			.553			.631		
ΔR^2				.163			.077		

Note. + p < .10, * p < .05, ** p < .01, *** p < .001

Moral Reasoning

In step 1, T2 criterion judgments were regressed on T1 criterion judgments, T1 EF, classroom anger/aggression, classroom prosocial behavior, and three interactions: T1 EF x T1 criterion, T1 EF x anger/aggression, and T1 EF x prosocial. T2 EF, T2 EF x anger/aggression, and T2 EF x prosocial were entered in step 2 (see Table 14). Classroom prosocial behaviors were entered in step 1 and negatively predicted criterion judgments, β = -.50, $t(31)$ = -3.25, p = .003. The T1 EF x prosocial interaction was entered in step 2 and negatively predicted criterion judgments, β = -.36, $t(30)$ = -2.54, p = .02. Follow up simple slopes test revealed that criterion judgments and EF were negatively associated for participants with high prosocial behaviors, $t(30)$ = -2.43, p = .02, but were positively associated for participants with low prosocial behaviors, $t(30)$ = 2.19, p = .04 (see Figure 7). Finally, classroom anger/aggression was entered in step 3 and negatively predicted criterion judgments, β = -.34, $t(29)$ = -2.46, p = .02. Prosocial behaviors and the T1 EF x prosocial interaction continued to predict T2 criterion judgments.

Table 14. T2 moral reasoning regressed on T1 moral reasoning, EF, temperament, and classroom behavior.

	Step 1			Step 2			Step 3		
Variables	*B*	*SE*	*β*	*B*	*SE*	*β*	*B*	*SE*	*β*
Prosocial Behaviors	-.07	.02	-.50**	-.07	.02	-.53***	-.09	.02	-.63***
T1 EF x Prosocial				-.03	.01	-.36*	-.03	.01	-.37**
Anger/Aggression							-.04	.02	-.34*
F	10.56**			9.42***			9.34***		
R^2	.254			.386			.492		
ΔR^2				.132			.106		

Note. * p < .05, ** p < .01, *** p < .001

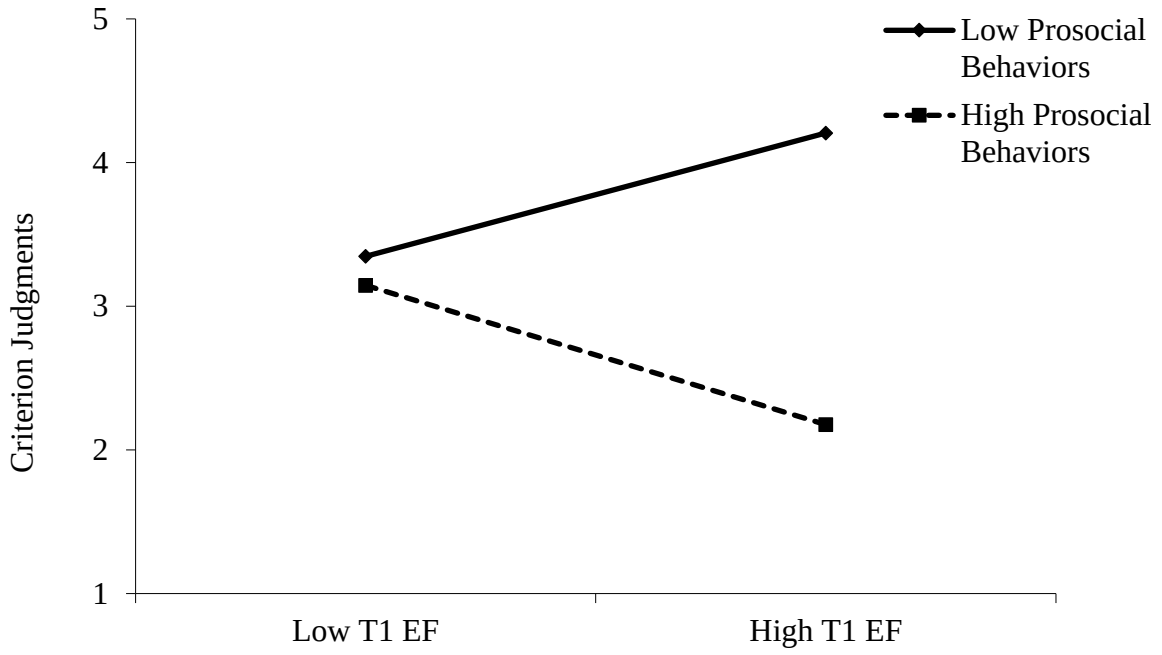

Figure 7. T2 criterion judgments regressed on T1 EF x prosocial

INTERIM SUMMARY: LONGITUDINAL RESULTS

T1 EF predicted T2 competent responses, but the association became insignificant with the entry of T2 EF. This suggests that T1 EF, as a main effect, has an indirect effect on T2 competent responses via its association with T2 EF. However, the T1 EF x surgency interaction emerged as a significant predictor of T2 competent responses, which suggests that T1 EF directly predicts T2 competent responses for participants high in temperamental surgency. Conversely, T1 competent responses did not emerge as a significant predictor of T2 EF, which suggests that the relation between EF and SC is asymmetrical. Surgency significantly predicted T2 EF, but this association became marginally significant when T2 competent responses were entered into the model. This suggests that T2 component responses accounts, in part, for the association between surgency and T2 EF.

Finally, T2 moral reasoning was predicted by both classroom prosocial behaviors and classroom displays of anger and aggression. Greater displays of anger and aggression predicted lower moral reasoning but interestingly, higher ratings of prosocial behaviors also predicted lower T2 moral reasoning. The significant interaction between T1 EF and prosocial behavior indicates that EF predicted moral reasoning, but a positive association between EF and moral reasoning was only observed in the context of low prosocial behavior. EF negatively predicted moral reasoning for children high in classroom displays of prosocial behavior.

Exploratory Analyses

Differences by Specific Response Type

All analyses up to this point have focused on the total number of competent responses chosen on the VSG, with the assumption that choosing any competent response over any aggressive response is indicative of SC. However, the competent response (i.e., avoid, tell the teacher, prosocial) may have differential relations with EF, temperament, and moral reasoning. Moreover, no attention has been given to the different types of aggressive responses. There is evidence to suggest that forms of overt aggression (e.g., hitting and yelling) may differ from each other (Bell & Willis, 2016), and that overt harm differs from relational aggression on multiple factors (e.g., Caporaso & Marcovitch, 2021; Crick et al., 2006; Crick & Warner, 1998; Ostrov et al., 2014; Murray-Close et al., 2006; Murray-Close et al., 2014), including EF relations during preschool (e.g., Poland et al., 2016). Therefore, the following exploratory analyses will look at the differential relations of EF, moral reasoning, temperament, and classroom variables for each of the six response choices, both at T1 with the full sample and longitudinally. Note that for moral reasoning, the following analyses use the average criterion variables for each form of retaliation (i.e., physical, verbal, and relational). Table 15 reports the frequencies of each response type at both time points.

Table 15. Frequencies of each response time at both time points.

T1 (*n* =86)						
# times response was chosen	Hit *n*	Yell *n*	Relational *N*	Avoid *N*	Tell the Teacher *n*	Prosocial *n*
0	43	55	47	51	36	7
1	31	22	29	27	37	15
2	8	8	7	6	7	15
3	3	1	3	1	4	21
4	0	0	0	0	1	12
5	1	0	0	0	0	11
6	0	0	0	1	1	5
T2 (*n* =33)						
# times response was chosen	Hit *n*	Yell *n*	Relational *N*	Avoid *N*	Tell the Teacher *n*	Prosocial *n*
0	28	28	25	22	9	2
1	3	3	7	10	13	1
2	2	1	1	1	9	5
3	0	1	0	0	0	7
4	0	0	0	0	1	7
5	0	0	0	0	1	4
6	0	0	0	0	0	7

Note. The numbers in the cells under each response option indicate the number of participants who chose that response type "x" number of times (e.g., at T1, n = 43 participants chose the hit response 0 times).

T1 Concurrent Analyses

Bivariate correlations for each of the six T1 response options and other T1 variables are presented in Table 16. Several notable trends emerged. First, EF was only correlated with hit responses, $r(86) = -.271$, $p = .01$, and prosocial responses, $r(86) = -.224$, $p = .04$, as a main effect, and with avoidant responses as an interaction with surgency, $r(86) = .212$, $p = .05$. Because the

EF x surgency interaction has a positive correlation with avoidant responses, similar to the EF x surgency interactions observed above, I speculate EF is only correlated with avoidant responses for participants with high surgency but not with low surgency.

Second, moral reasoning was primarily related to hit responses as a main effect, $r(86)_{hitcriterion} = -.339$, $p = .001$, and as an interaction with surgency, $r(86)_{hitxsurgency} = -.261$, $p = .02$. The criterion judgments for yelling and relational aggression did not relate to the respective response options either as a main effect or as part of an interaction, all *p*s > .05. There was not a compelling pattern that emerged between moral reasoning and any particular competent response, although a few significant interactions with moral reasoning were observed with avoidant and prosocial responses. Thus, the observed association between criterion judgments and competent responses at T1 indicates that moral reasoning is associated with avoiding aggressive responses (particularly physical aggression) but does not necessarily support competent responding in of itself.

Third, associations between temperament, both as main effects and with interactions with other variables, were observed at least once for all response options except for the tell the teacher response. Notably, surgency was negatively associated with prosocial responses, $r(86) = -.267$, $p = .01$. The gender x negative affect interaction observed in the T1 concurrent responses primary regression analysis was only correlated with prosocial responses, $r(86) = .398$, $p < .001$, likely reflecting the same pattern seen in Figure 3 (i.e., negative affect is positively associated with prosocial responses for girls, but negatively associated with prosocial responses for boys).

Fourth, gender was implicated in a few associations with relational aggressive responses and moral judgments about relational retaliation. Specifically, gender x surgency is associated with relational aggressive responses, $r(86) = -.280$, $p = .01$, and relational criterion judgments x gender are associated with prosocial responses, $r(86) = -.230$, $p = .03$. Additional examination of these associations is needed to understand the nature of the interactions, but they are notable given potential gender differences in displays of relational aggression during preschool (e.g., Crick et al., 2006; Ostrov et al., 2014; Poland et al., 2016).

Finally, the data suggest that overall trends in the primary T1 concurrent regression analysis were mostly driven by the hit response and the prosocial response, as well as the avoidant response. In the T1 concurrent regression analysis, EF, criterion judgments, and the gender x negative affect interactions all emerged as significant, independent predictors of T1 competent responses. The hit response and the prosocial response were each associated with two of these predictors and the avoidant response was associated with one (hit criterion judgments).

Table 16. Correlations between T1 VSG responses, EF, moral reasoning, temperament, and gender.

	Hit	Yell	Relational	Avoid	Tell the Teacher	Prosocial
EF	-.271*	-.184	-.058	.049	.008	.224*
Physical Criterion	-.339**	.010	-.186	.253*	-.117	.206
Verbal Criterion	-.150	-.184	.070	-.042	.037	.129
Relational Criterion	-.337**	-.028	.091	.093	.009	.098
Surgency	.185	.154	-.031	.122	.075	-.267*
Negative Affect	.042	-.075	.065	.023	.004	-.037
EF x Surgency	-.034	-.091	.012	.212*	.036	-.085
EF x Negative Affect	-.102	.028	-.135	.121	.096	-.017
Gender x EF	-.103	-.008	-.007	-.120	.099	.067
Gender x Surgency	.087	.017	.280**	.094	-.051	-.207
Gender x Negative Affect	-.072	-.180	-.174	-.165	-.176	.398**
EF x Physical Criterion	.164	-.128	-.026	.028	-.040	-.014
EF x Verbal Criterion	.084	-.037	.090	-.117	.078	-.058
EF x Relational Criterion	.202	-.105	.005	-.113	-.050	.023
Physical Criterion x Surgency	-.261*	.106	-.018	.073	.124	-.010
Verbal Criterion x Surgency	-.008	-.067	.016	.147	.066	-.095
Relational Criterion x Surgency	-.106	-.041	.152	.078	.164	-.139
Physical Criterion x Negative Affect	-.067	-.132	-.034	.132	-.046	.067
Verbal Criterion x Negative Affect	-.097	.022	-.038	.213*	-.144	.037
Relational Criterion x Negative Affect	-.148	-.307**	-.125	.093	.125	.145
Physical Criterion x Gender	-.021	-.017	-.096	-.140	-.093	.198
Verbal Criterion x Gender	-.027	-.118	-.117	-.201	.069	.187
Relational Criterion x Gender	-.078	-.056	-.082	-.056	-.151	.230*

Note. *p < .05, ** p < .01. Marginal significance (p < .10) was not considered for exploratory analyses.

Longitudinal Analyses

Bivariate correlations were run between each of the six T2 response options and the T1 predictor variables (see Table 17). Many of the trends observed at T1 were also observed at T2. Specifically, EF and criterion judgments were associated with the hit response, the criterion judgments specific to yelling and relational aggression were not correlated with their respective responses, and gender was implicated in relational aggression responses via interactions with the temperament variables.

There were a few additional patterns that emerged that provide insight into the specific responses that contributed to the overall findings observed in the primary longitudinal analysis (i.e., T1 EF and surgency indirectly and T1 EF x surgency directly predicted T2 competent responses). First, T1 EF was only correlated with prosocial responses via the T1 EF x surgency interaction, $r(33) = .386$, $p = .03$. As a main effect, EF was negatively associated with both hit, $r(33) = -.446$, $p = .01$, and yell, $r(33) = -.379$, $p = .03$, responses. This suggests that the indirect effect of T1 EF was not related to the ability to pick a competent response at T2, but rather to avoid picking an aggressive response. The T1 EF x surgency interaction was also associated with the yell response, $r(33) = -.565$, $p < .001$, and the main effect of surgency was associated with both yell, $r(33) = .560$, $p < .001$, and prosocial, $r(33) = -.466$, $p = .01$, responses. Thus, the longitudinal analysis results appear to be mostly driven by the hit, yell, and prosocial responses.

Table 17. Correlations between T2 VSG responses, T1 EF, T1 moral reasoning, temperament, and gender.

	Hit	Yell	Relational	Avoid	Tell the Teacher	Prosocial
EF	-.446**	-.379*	-.120	.101	-.030	.308
Physical Criterion	-.436*	.037	-.048	-.272	-.398*	.483**
Verbal Criterion	-.492**	-.233	.002	.012	-.002	.327
Relational Criterion	-.451**	.088	-.198	.160	-.323	.240
Surgency	.052	.560**	.196	.288	.134	-.466**
Negative Affect	.198	-.232	-.251	-.270	-.005	.189
EF x Surgency	.075	-.565**	-.217	-.181	-.111	.386*
EF x Negative Affect	-.325	.417*	-.106	.396*	-.277	.031
Gender x EF	-.082	-.109	.119	.038	.110	-.052
Gender x Surgency	-.145	.069	-.373*	.303	-.155	.136
Gender x Negative Affect	.042	-.211	.371*	-.059	-.136	.065
EF x Physical Criterion	.410*	-.053	-.034	-.006	.025	-.113
EF x Verbal Criterion	.422*	.090	-.242	.272	-.006	-.176
EF x Relational Criterion	.389*	-.325	-.160	-.096	-.042	.107
Physical Criterion x Surgency	-.312	.140	.217	.202	-.181	.034
Verbal Criterion x Surgency	-.034	-.268	.159	-.107	.020	.086
Relational Criterion x Surgency	-.288	.168	.186	.381*	-.154	.048
Physical Criterion x Negative Affect	-.263	.158	.095	.094	-.264	.137
Verbal Criterion x Negative Affect	-.280	.259	.114	.076	-.132	.018
Relational Criterion x Negative Affect	-.279	-.199	.158	-.119	-.243	.313
Physical Criterion x Gender	.237	-.101	-.166	.117	.304	-.223
Verbal Criterion x Gender	.118	-.087	.003	.142	.067	-.093
Relational Criterion x Gender	.189	.271	-.188	.150	.109	-.225

Note. *p < .05, ** p < .01. Marginal significance (p < .10) was not considered for exploratory analyses.

Differences in Moral Reasoning by Retaliation Type and Judgments

According to social domain theory, children with mature moral concepts recognize that harm is wrong independent of both authority and rules (i.e., criterion judgments; Smetana et al., 2014). On measures of moral reasoning, children with mature moral concepts consequently rate transgressions as equally "bad" across the acceptability and criterion judgments. In addition, children rate unprovoked physical harm (e.g., hitting) as more morally wrong than acts of unprovoked psychological harm (e.g., relational or verbal aggression; Murray-Close et al., 2006; Smetana et al., 2003; Smetana et al., 1999). Therefore, it is possible that preschool children may have a more mature moral concept for the hit retaliation compared to the yelling and relational retaliation.

A 2 (gender) x 2 (EF group) x 3 (retaliation type) x 4 (judgment type) mixed ANOVA was run on the children's retaliation judgments. Gender was included as a between-subjects variable based on differences in observed levels of aggression (e.g., Crick et al., 2006; Ostrov et al., 2014; Poland et al., 2016). Perhaps these differences are reflective of differences in moral judgments as well. EF was also included as previous analyses have demonstrated that EF is associated with criterion judgments. Participants were grouped as either high or low EF based on a mean split (i.e., participants with a composite EF score over 0 were considered "high EF" and participants with a composite score less than 0 were considered "low EF").

The mixed ANOVA revealed a main effect of judgment, $F(3, 82) = 28.09$, $p < .001$, $\eta^2_p = .26$ (see Figure 8). Posthoc LSD pairwise comparisons revealed that acceptability judgments ($M = 4.25$, $SD = 0.94$) were significantly higher than authority independence ($M = 3.67$, $SD = 1.04$) and rule independence ($M = 3.56$, $SD = 1.15$) criterion judgments, ps $< .001$, which in turn were both higher than deserved punishment judgments ($M = 3.04$, $SD = 1.08$), ps $< .001$. The two criterion judgments did not differ from one another, $p = .39$. No interactions with judgment were significant, all ps $> .10$. The lack of an interaction between retaliation type and judgment, $F(6, 82) = 0.32$, $p = .93$, suggests that the participants did not have a more mature moral concept for physical retaliation over the two forms of psychological retaliation. If they had, the criterion

judgments for physical retaliation would have been significantly higher than the criterion judgments for both verbal and relational retaliation.

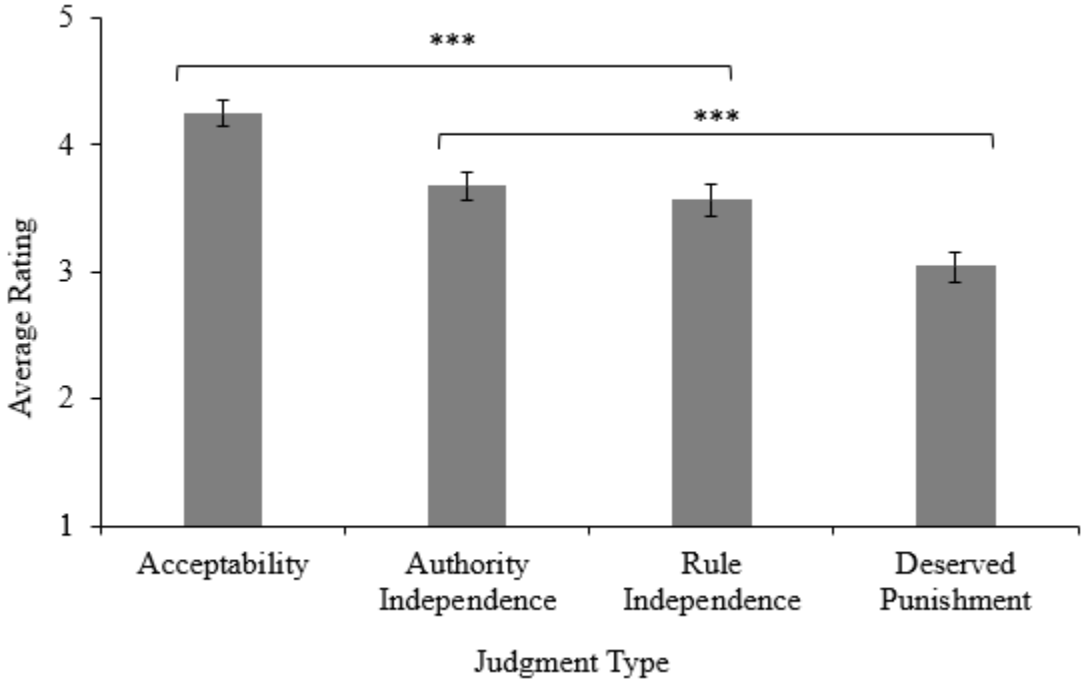

Figure 8. Main effect of judgment type for moral reasoning. *Note*. *** p < .001

There was no main effect of retaliation type, $F(2, 82) = 0.99$, $p = .37$. However, there was a significant interaction between retaliation type and gender, $F(2, 82) = 3.70$, $p = .03$, $\eta^2_p = .04$ (see Figure 9). A follow up 2 (EF group) x 3 (retaliation type) x 4 (judgment type) mixed ANOVA split by gender revealed that boys' moral reasoning differed by retaliation type, $F(2, 38) = 3.68$, $p = .03$, $\eta^2_p = .09$, but girls' moral reasoning did not differ by retaliation type, $F(2, 44) = 1.21$, $p = .31$. Posthoc LSD pairwise comparisons revealed that boys' physical retaliation judgments ($M = 3.83$, $SD = 0.99$) were significantly higher than their verbal retaliation judgments ($M = 3.49$, $SD = 1.08$), $p = .04$, and relational retaliation judgments ($M = 3.50$, $SD = 1.09$), $p = .01$. Judgments for verbal and relational retaliation did not differ, $p = .95$. Further exploratory independent samples t-tests indicated that girls' and boys' judgments did not differ for any retaliation types, all ps > .10.

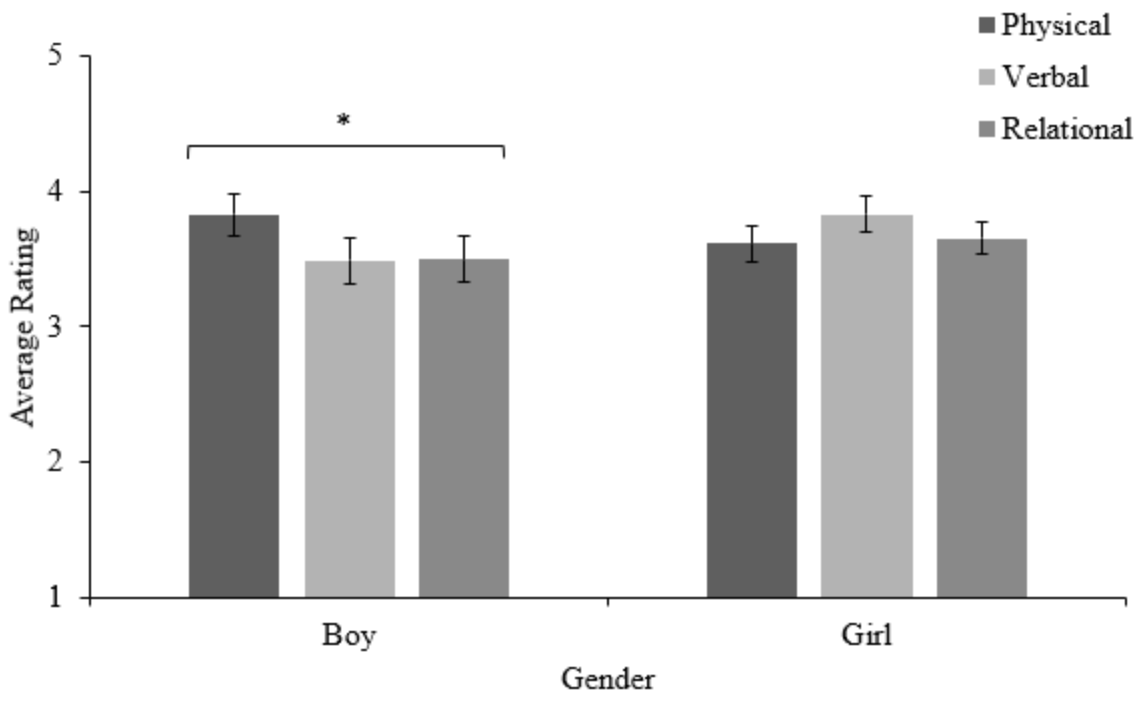

Figure 9. Gender x retaliation type interaction for moral reasoning. Note. * p < .05

There was also a marginally significant interaction between retaliation type and EF group, F (2, 82) = 2.89, p = .06, η^2_p = .03 (see Figure 10). A follow up 2 (gender) x 3 (retaliation type) x 4 (judgment type) mixed ANOVA split by EF group revealed that low EF participants' moral reasoning differed by retaliation type, F(2, 38) = 3.74, p = .03, η^2_p = .09, but the high EF participants' moral reasoning did not differ by retaliation type, F(2, 44) = 0.39, p = .68. Posthoc LSD pairwise comparisons revealed low EF participants' physical retaliation judgments (M = 3.75, SD = 1.03) were significantly higher than their judgments of relational retaliation (M = 3.34, SD = 1.03), p = .003. Verbal aggression judgments (M = 3.50, SD = 1.08) did not differ from either physical or relational retaliation, both ps > .10. Further exploratory independent samples t-tests indicated that high EF participants' relational retaliation judgments (M = 3.79, SD = 0.90) were higher than low EF participants' relational retaliation judgments, t(84) = 2.01, p = .05.

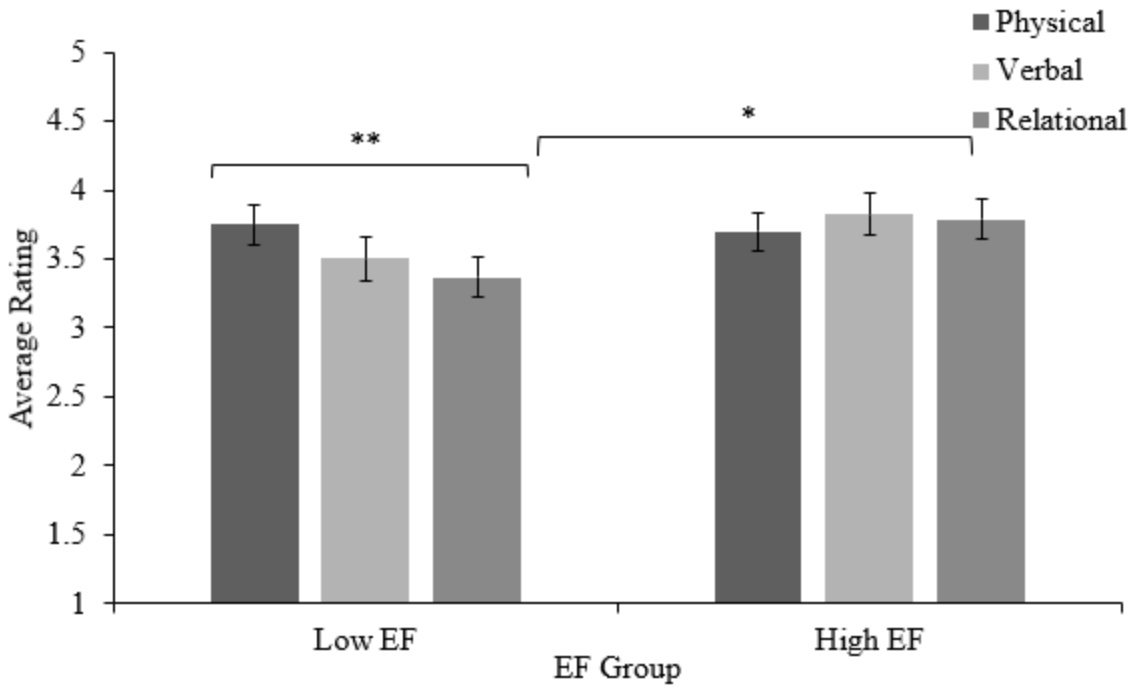

Figure 10. EF group x retaliation type interaction for moral reasoning. Note. * p < .05, ** p < .01

Finally, both of these interactions were further qualified by a marginally significant gender x EF group x retaliation type three-way interaction, $F(2, 82) = 2.91$, $p = .06$, $\eta^2_p = .03$ (see Figure 11). A follow up 2 (EF group) x 3 (retaliation type) x 4 (judgment type) mixed ANOVA split by gender revealed that the EF group x retaliation type interaction was significant for girl participants, $F(2, 44) = 5.23$, $p = .01$, $\eta^2_p = .11$, but not for boy participants, $F(2, 38) = 0.34$, $p = .01$. A follow up 3 (retaliation type) x 4 (judgment type) repeated measures ANOVA split by EF group was run for girl participants. There was a significant main effect of retaliation type for both low, $F(2, 16) = 3.27$, $p = .05$, $\eta^2_p = .17$, and high EF girls, $F(2, 16) = 4.28$, $p = .02$, $\eta^2_p = .13$, but posthoc LSD pairwise comparisons revealed that the effect is in the opposite direction for each EF group. Low EF girls' judgments of relational aggression ($M = 3.46$, $SD = 0.87$) were significantly lower than their judgments of physical aggression ($M = 3.90$, $SD = 0.83$), $p = .04$, and marginally lower than their judgments of verbal aggression ($M = 3.85$, $SD = 0.84$), $p = .08$. High EF girls' judgments of physical aggression ($M = 3.34$, $SD = 0.91$) were significantly lower

than both their judgments of relational (M = 3.83, SD = 0.75) and verbal (M = 3.85, SD = 0.87) aggression, both ps = .02.

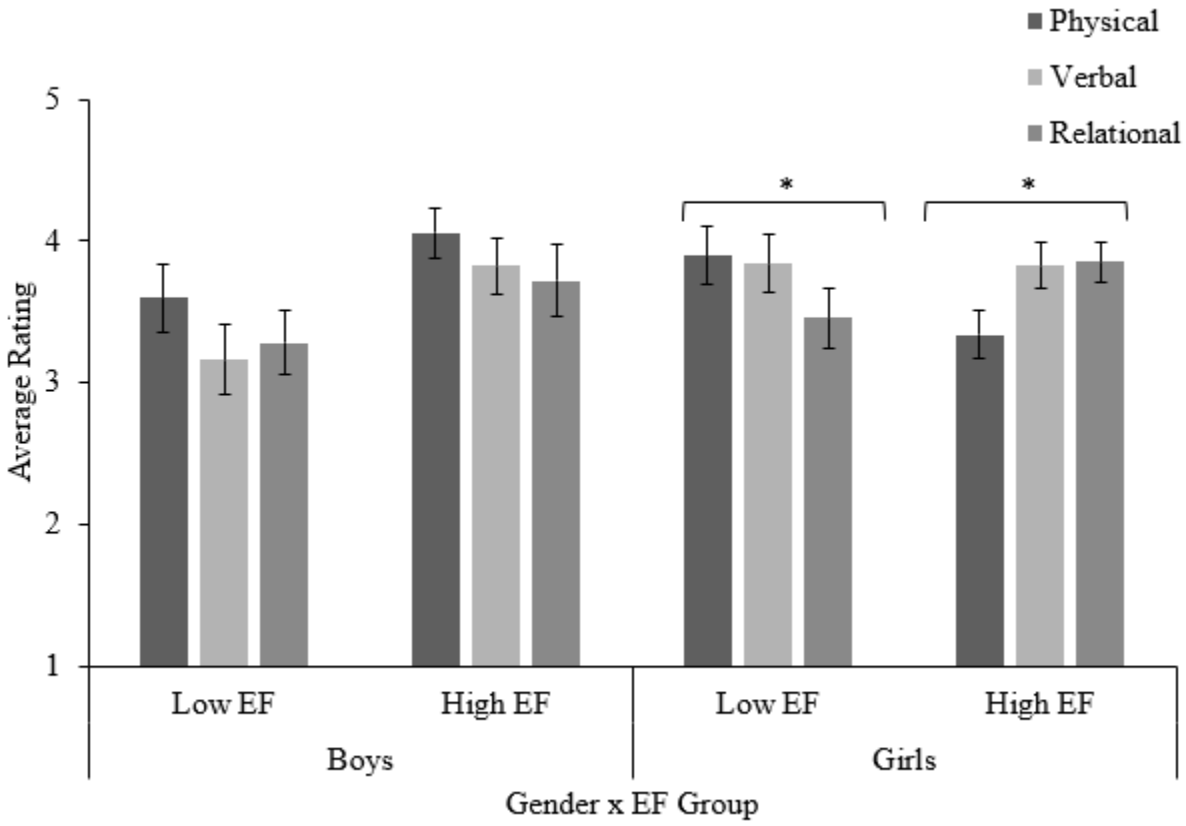

Figure 11. Gender x EF group x retaliation type interaction for moral reasoning. *Note.* * $p < .05$

TEACHER RANKINGS OF RESPONSE OPTIONS

As part of the teacher questionnaires, teachers were asked to rank the appropriateness of each of the six response types on the VSG (1 = most appropriate, 6 = least appropriate). The assumption was that prosocial responses would be the most appropriate and hit would be the least appropriate, yet it was less clear how teachers would rank the remaining competent (tell the teacher and avoid) and aggressive (yell and relational aggression) responses. Teachers ranked the six response choices overall, and then were provided with four different unprovoked transgression scenarios (i.e., physical aggression transgression, teasing transgression, social

exclusion transgression, and resource transgression) and were asked to rank the appropriateness of each response within the context of the scenarios. The survey was completed by 16 teachers, but 2 teachers skipped the "overall ranking" question.

The frequency of overall rankings for each response type is listed in Table 18. It is clear from the ranking frequencies that teachers differentiated competent from aggressive responses in the same way the responses were grouped in the current study. Across the 14 teachers, none of the competent responses received a ranking below 3 and none of the aggressive responses received a ranking above 4. In addition, most teachers ranked the prosocial response as the most appropriate and the hit response as the least appropriate. A repeated measures one-way ANOVA with posthoc LSD tests at the .05 level confirmed that teachers ranked prosocial responses as the most appropriate response, followed by telling the teacher and avoidant responses (which did not significantly differ from one another), followed by yell and relational aggression (which did not significantly differ from one another), and finally hit was ranked as the least appropriate response, $F(5, 13) = 134.65$, $p < .001$, $\eta^2_p = .91$.

A series of repeated measures one-way ANOVAs were then conducted to see whether teachers differentially rated the response types based on the transgression type. Except for the tell the teacher response, teachers did not differentially rank any of the response types based on the transgression type, all *p*s > .05. Teachers' rankings of the tell the teacher response differed by the transgression type, $F(5, 15) = 4.16$, $p = .01$, $\eta^2_p = .22$. Posthoc LSD tests at the .05 level revealed that teachers ranked the tell the teacher response as more appropriate in response to physical aggression ($M = 2.19$, $SD = .66$) compared to social exclusion ($M = 2.81$, $SD = .75$) and marginally more appropriate than in response to teasing ($M = 2.56$, $SD = .63$). In addition, teachers ranked the tell the teacher response as more appropriate in response to resource transgressions (i.e., a toy is taken away; $M = 2.31$, $SD = .48$) compared to social exclusion.

The results from this survey provide external validation for how responses were grouped in the current study. In addition, these results provide evidence that teachers agree on the most appropriate and least appropriate response types to conflict and that they mostly do not differentiate based on the type of unprovoked transgression. The differentiation for the tell the

teacher response suggests that teachers may see that some transgressions need more adult intervention (e.g., physical aggression transgressions) than other transgressions (e.g., social exclusion transgressions). Perhaps teachers believe that relational issues are issues that children should generally try to handle themselves. However, more research is required with a larger sample size of teachers from diverse childcare centers to investigate this claim further.

Table 18. Frequencies of overall rankings for each response types

Ranking	Hit	Yell	Relational	Avoid	Teacher	Prosocial
1 (most appropriate)	0	0	0	0	2	12
2	0	0	0	8	4	2
3	0	0	0	6	8	0
4	0	5	9	0	0	0
5	1	8	5	0	0	0
6 (least appropriate	13	1	0	0	0	0

Interim Summary: Exploratory Analyses

The exploratory analyses revealed several notable findings regarding the specific response options on the VSG, moral reasoning, and teachers' views of conflict response options. First, the correlation analyses demonstrated differential associations between the predictor variables and the six response options on the VSG. Based on these associations, the primary analysis results were mostly driven by the hit, avoidant, and prosocial response choices at T1 and the hit, yell, and prosocial response choices longitudinally at T2. In addition, moral reasoning was negatively related to the hit response choice, but not the yell or relational aggression response choices at T1 and longitudinally at T2.

Second, the moral reasoning exploratory analyses revealed that preschool children do not have a mature moral concept of retaliation, as demonstrated by the significant difference between the acceptability and criterion judgments for all three types of retaliation. As a group, boys had the highest moral reasoning ability for physical aggression compared to verbal or relational

aggression. Girls, however, had differential patterns based on their level of EF. Low EF girls had lower moral reasoning for relational aggression compared to physical and verbal aggression whereas high EF girls had lower moral reasoning for physical aggression compared to verbal and relational aggression.

Finally, the teacher ratings of the VSG response options demonstrate that preschool teachers see the prosocial response as the most appropriate response to conflict and the hit response as the least appropriate response to conflict, followed by the yell and relational response types, which were ranked as less appropriate than the avoidant and tell the teacher response types. For the most part, these rankings did not differ by the type of conflict (e.g., a physical aggression conflict vs. a resource conflict).

CHAPTER IV: DISCUSSION

The primary goal of the current study was to examine how EF contributes to the development of SC in peer conflict situations during children's final year of preschool. The results indicate that EF is longitudinally associated with SC, but importantly, the direct effect of EF was only for participants high in temperamental surgency. The secondary goal of the current study was to examine the developmental trends of moral reasoning regarding retaliation. Children's classroom behavior predicted moral reasoning at the end of preschool, but there is some evidence to suggest that EF is also an important factor for moral reasoning development. In service of the study's goals, many additional analyses were conducted to provide an in-depth account of both SC and moral reasoning development during the preschool years.

The Role of EF in the Development of SC

EF was of particular interest given its robust association with SC during preschool (e.g., Denham et al., 2014; O'Toole et al., 2019; Poland et al., 2016; Rathert et al., 2011), its marked improvement during preschool (e.g., Alink et al., 2006; Carbonneau et al., 2015), its role in age-related changes of SC (Caporaso et al., 2019), and its longitudinal associations with aggressive behavior (e.g., Kochanska & Knaack, 2003; Nigg et al., 1999). The results of the current study are consistent with these previous findings and extend this line of research through several novel contributions. The longitudinal results provide insight into the mechanisms through which EF supports SC development. Borrowing from the theory of mind literature (e.g., Moses, 2001), it was proposed that EF's association with SC could reflect one of two accounts: an *expression* account or an *emergence* account. The *expression* account suggests that EF provides in-the-moment control as children face provocation. EF essentially assists children with meeting the "task demands" of social conflict situations. As children's EF develops, so does their ability to control their behavior during social conflict.

Conversely, the *emergence* account suggests that EF is a necessary, foundational ability for the emergence of a SC concept. Early EF abilities provide children with the experiences necessary for developing the social scripts and schema necessary for competence in social conflict (e.g., Crick & Dodge, 1994). The emergence account implies a cycle in which low EF children are more likely to use aggression in early conflict situations, leading to the development of maladaptive beliefs, response biases, and concepts, which in turn influence continued aggressive behavior.

In the theory of mind literature, studies that show that early EF directly predicts later theory of mind and that early theory of mind does not predict EF are considered evidence for an emergence account (Carlson et al., 2015; Marcovitch et al., 2015). Evidence for an expression account would include an indirect effect of early EF via its association with later EF, or the presence of a symmetrical relation (i.e., early theory of mind predicts later EF). Based on these interpretations, the longitudinal analyses in the current study provide evidence for both emergence and expression accounts. Although there was little evidence for a symmetrical relation between EF and SC, T1 EF only had an indirect association with T2 competent responses via T2 EF. This suggests, at least for some children, that current levels of EF are more important for competence in conflict situations than prior EF. Children with low EF at T1 are more likely to have low EF at T2 but having low EF at T1 does not automatically place children on a high-risk path for aggression issues due to issues with a maladaptive SC concept. However, the significant T1 EF x surgency interaction suggests that this is not true for all children.

Children high in temperamental surgency, characterized by high approach tendencies and impulsivity in social situations (Rothbart & Bates, 2006), may be at risk for developing maladaptive SC concepts if they also have low EF. Due to their sociability and impulsivity, exuberant children may be more likely to be involved in peer conflict (Berdan et al., 2008; Dollar & Stifter, 2012; Mendez et al., 2002). Exuberant children with low EF are then more likely to respond to conflict with aggression. Because these situations happen with a greater frequency for low EF exuberant children, these children may be deprived of the peer interactions that would foster the development of social skills and they may begin to develop maladaptive concepts and beliefs regarding aggression (O'Toole et al., 2019). For example, continued use of aggression

may strengthen mental representations of aggressive response options such that they come to mind more readily during conflict (Crick & Dodge, 1994). In addition, these children may begin to believe that they are better at enacting aggressive responses and are not effective at enacting competent responses (Erdley & Asher, 1996). As these maladaptive concepts develop and become further ingrained, it may become harder for children to override aggressive tendencies in favor of competent responses. Thus, EF may be particularly important for children high in surgency to prevent them from beginning this cycle.

These findings further suggest that EF may not be necessary for all children to behave competently when faced with provocation, particularly by the end of preschool. Children with low T1 EF and low surgency had equally high T2 competent responses as low surgent children with high EF. The EF x surgency interaction was also observed concurrently at T2. However, this interaction was not observed in the T1 concurrent data for either the full sample or the longitudinal subsample. In fact, there were no significant interactions between EF and any other variable at T1. One possible explanation is that aggressive behavior may still be a rather normative experience at the beginning of preschool, particularly for children who are on the decreasing trajectory for aggressive behavior (Alink et al., 2006). These children show minimal levels of aggression by the time they are 5, the age that most children turn during the preschool year (Carbonneau et al., 2016; NICHD Early Child Research Network, 2004). Perhaps, then, the EF main effect observed at T1 also captures children who will otherwise be non-aggressive by the time they reach the end of preschool. By the end of the year, the children who need to rely on EF to control aggressive behavior are those who are high in surgency.

However, the lack of a T1 EF interaction does not definitively show that EF is necessary for SC for all children before 5 years of age. There are other variables that were not measured in the current study that could potentially affect the relation between EF and SC. Time spent in daycare settings is one primary factor that has been found to differentiate members of low aggression trajectories compared to the moderate or high trajectory (NICHD Early Child Care Research Network, 2004). Another factor is the presence of older siblings in the home at the time of birth (Tremblay et al., 2004). In fact, the presence of an older sibling in the home was identified as the largest risk factor of being in the high aggression group, and it was the only factor among a

variety of SES indices and parenting behaviors that differentiated between the low and moderate trajectory. Both of these factors highlight the significance of having similar-aged peers in the immediate environment for the development of aggressive tendencies. The more time children spend around their peers, the more likely they are to get into conflict with those peers, and the more opportunity to respond to conflict with aggression. Tremblay et al. further suggest that younger siblings may be less likely to be punished for aggression, which could reinforce the notion that aggression is an appropriate way to resolve conflict. By contrast, children who spend less time around peers have fewer opportunities to use aggressive responses during conflict, and when if they do use aggression with their parents, they may be more likely to be reprimanded. Consequently, aggressive responses may not come to them as readily in the moment of peer conflict, so they do not need to rely on EF to respond competently.

Yet the differential associations observed between EF and specific response choices indicate that the different types of aggressive responses (as well as competent responses) have differing EF demands. EF was only associated with hit and prosocial responses at T1 and hit and yell responses at T2. T2 prosocial and yell responses were associated with the EF x surgency interaction. This interaction presumably indicates that EF was only associated with these responses for participants high in surgency given the pattern observed with T2 competent responses.

Similar findings were found by Caporaso and Marcovitch (2021). In this study, participants who took part in a cognitively taxing procedure chose less aggressive and more prosocial responses, but the procedure did not have an effect on the yell or avoid responses. The authors concluded that children may view yelling as a less serious offense compared to physical aggression so they may not use EF to inhibit choosing a yell response by comparison. The avoid response is a script-based response (as is the tell the teacher response) that can be applied to most provocation situations, but prosocial responses have to be tailored to the specific situation. Even if the response choices are provided for the participants, such as in both Caporaso and Marcovitch and the current study, participants may still need a certain level of EF to recognize that prosocial choices are the most optimal compared to the other competent responses. Indeed, these assumptions are corroborated by the teacher ratings of the response types. Teachers almost

unanimously rated physical aggression as the worst response and prosocial responses as the best response to provocation. These beliefs are likely translated to the messages they tell their students when talking about how to handle instances of conflict that arise in the classroom.

Interestingly, the yell response was only associated with EF at T2. Perhaps this is because prior to pre-kindergarten, children are frequently told not to be physically aggressive but to instead "use their words". Young children are consistently reminded that physical aggression is wrong but may receive less feedback when they engage in verbal aggression. However, once acts of physical aggression begin to decrease, authority figures may focus on more nuanced social problem-solving behavior, such as using a nice tone of voice and saying nice words. By the end of preschool, children may begin to use EF to also avoid forms of verbal aggression per these messages. However, it is worth noting the somewhat contradictory finding with the yell response results in Caporaso and Marcovitch (2021), particularly given that participants in that study were the same age (if not slightly older) as the participants at T2.

Finally, the individual EF tasks had differential associations with competent responses at both time points and longitudinally. Concurrently, H/S Stroop was the only EF task related to competent responses at T1, while it was the only EF task *not* related to competent responses at T2. Despite the fact that T1 DCCS and T1 MST were not associated with T1 competent responses, both were associated with T2 competent responses (and presumably account for the observed indirect longitudinal EF effect). These results are at odds with previous research, which typically finds correlations between all three EF components measured in the current study and SC (Caporaso et al., 2019; Poland et al., 2016). In addition, studies that have assessed the relative contributions of EF component have found that working memory tends to be the only component that independently predicts SC (Caporaso et al., 2019; Caporaso et al., 2021), so it was unexpected that the T1 MST was not associated with T1 competent responses.

Although the study was not designed to examine the differences with each EF component, these findings are nonetheless noteworthy and should be explored in further research. I speculate that the observed differential associations could reflect developmental changes in the utility of EF in SC as both constructs develop during preschool. Perhaps inhibition is more important as children

are still learning how to respond appropriately to conflict and may still be drawn to aggressive options. As children further develop and practice competent responses, they may need to rely less on inhibition processes, but more so on working memory and cognitive flexibility to construct a SC concept, bring those response options to mind, and recognize that they are the most optimal responses among an array of response choices. Indeed, inhibition becomes less important for SC with age. In a similar longitudinal design, O'Toole et al. (2019) found that inhibition was associated with physical aggression at 5 years of age but was no longer significant at 5.5 and 6 years of age. This pattern appears to persist over the course of childhood. In a sample of 9-year-old children, Granvald and Marciszko (2016) found that working memory and cognitive flexibility were associated with reactive aggression, but inhibition was not.

Other Contributions of Temperament

Surgency emerged as a particularly important temperamental factor for the development of SC over the preschool year. Surgency also negatively predicted T2 EF, but the association was partly accounted for by T2 competent responses. This finding is consistent with a mediation model but additional research with a much larger sample is needed to test this assertion. Despite the importance of surgency for the T2 outcome measures, negative affect appeared to be the temperamental dimension central to SC at the beginning of preschool. Although a few T1 associations were observed between surgency (as either a main effect or interaction) and competent response variables, surgency did not independently predict T1 competent responses in the full sample or the longitudinal subsample. Negative affect, however, predicted T1 competent responses in the longitudinal subsample as a significant main effect and as a marginal interaction with criterion judgments. These results are generally consistent with research that has found associations between negative affect and social behavior during preschool (Cohen & Mendez, 2009; Eisenberg et al., 2009; Gilliom et al., 2002; Mendez et al., 2002; Moran et al., 2013; Zeman et al., 2002), theories that implicate anger in aggressive responses to provocation (Lemerise & Arsenio, 2000; Orobio de Castro, 2004), and research that has found that children's projected feelings of anger in a hypothetical situation are associated with subsequent aggressive response choices (Denham et al., 2014).

In the full sample, the negative affect x gender interaction predicted T1 competent responses. Interestingly, the expected direction of this effect was only observed for boys, while the opposite effect was observed for girls. High negative affect boys were less likely to choose competent responses to conflict but high negative affect girls were more likely to choose competent responses to conflict. Although much of the research on negative affect and aggression fails to find any difference by gender (e.g., Eisenberg et al., 2009; Moran et al., 2013; Zeman et al., 2002), Eisenberg et al. (1995) found that negative affect was associated with SC for only boys in a sample of 5- to 7-year-old children across multiple time points, with multiple different measures of negative affect and social behavior, in both the school and home contexts. Yet no research to date has found that negative affect supports SC for girls. These results cannot be attributed to simple gender differences in negative affect, as no gender difference was observed in parent ratings of negative affect in the current study, a finding consistent with previous research (for review, see Gagne et al., 2013).

Perhaps parental emotion socialization practices could explain the observed pattern of results. In general, punitive parental responses to negative emotion are related to lower ratings of child SC (Eisenberg et al., 1996; Fabes et al., 2001) and supportive responses are related to higher SC (Eisenberg et al., 1996; Thompson et al., 2020). Differences in parental emotion socialization practices based on child gender are not observed in recent research with United States samples (Denham et al., 2010). However, the same socialization practices appear to have a differential effect on boys and girls in relation to SC (Eisenberg et al., 1996; Jones et al., 2002; but see Thompson et al., 2020). Girls' socioemotional competence is more negatively affected by punitive or minimizing socialization than boys, particularly if they are high in negative affect (Jones et al., 2002). In addition, positive socialization has a greater impact on girls' emotional knowledge than boys' (Denham et al., 2010). As such, it is thought that girls may be more susceptible to socialization messages (Denham et al., 2010). In a higher SES sample, such as the one in the current study, parents tend to use more supportive rather than dismissive socialization strategies (Pinderhouse et al., 2000). The high negative affect girls in the current study may have received more emotion socialization from their parents in response to their frequent displays of negative emotions, giving them more opportunities to learn about emotions and how to handle

them appropriately. Emotion knowledge is then associated with both teacher and parent-reported SC (Thompson et al., 2020), as well as competent choices to hypothetical social situations (Denham et al., 2013).

Boys, on the other hand, may not be able to take advantage of these instances of socialization or be able to apply the messages in the context of social conflict as well as girls. In addition, they are more likely to use avoidant strategies when experiencing negative emotions whereas girls are more likely to approach their parents and verbally express their discomfort (Denham et al., 2010; Permo & Kiel, 2014). Parents then have less opportunity to engage in emotion socialization and are more likely to respond in a punitive manner when their sons use avoidant coping strategies (Permo & Kiel, 2014). Paired with the finding that punitive responses are associated with lower emotional knowledge and worse SC, it is perhaps not a surprise that boys that experience frequent and intense negative emotions are more likely to choose aggressive responses to conflict. However, these pathways are all based on speculation and need to be empirically tested in future research.

The gender x negative affect interaction did not predict T2 competent responses either concurrently or longitudinally. This is not particularly noteworthy given that the interaction was only significant for the full T1 sample and was not observed in the T1 longitudinal subsample. Yet it is interesting that negative affect did not predict T2 competent responses despite concurrent T1 associations in the longitudinal subsample. In fact, none of the zero-order associations between T2 competent responses and negative affect or any of its interactions were significant. These results are particularly surprising given the research that has shown that infant negative affect ("difficult temperament") predicts kindergarten aggressive behavior (Vitaro et al., 2006).

The reason for the apparent trade-off between negative affect and surgency over the span of the pre-kindergarten year is unknown. I speculate that this could be due to changes in negative affect, particularly anger, over time. Even though temperament is considered to be stable over time, normative increases and decreases in anger are observed throughout childhood and into adolescence (Cole et al., 2011). Displays of anger, just like aggression, increase during

toddlerhood and then decrease through the preschool years (Cole et al., 2011). Parents filled out the CBQ-VSF at T1 and it is possible that their ratings captured normative displays of anger, as well as stable, trait anger. But as self-regulatory processes develop during the preschool year (including EF), children become better at regulating anger (Dollar & Calkins, 2019). When paired with pre-kindergarten curricula that tend to emphasize socioemotional knowledge, anger may not be as much of a factor for children who are otherwise low in trait anger by the end of preschool. Measures of negative affect during infancy may be a more accurate indicator of a stable predisposition towards anger because it is measured prior to the normative increase in the toddler years.

Moral Reasoning

The results from the current study contribute several novel findings regarding preschool children's moral reasoning. Most importantly, the current study provides evidence that children's moral reasoning about retaliation is positively related to their responses to provocation. At T1, participants who had a more mature moral concept of retaliation, indicated by higher ratings on the criterion judgment questions, chose more competent responses to peer conflict. This finding is in line with previous work that has found that children's moral knowledge and beliefs relate to their moral behavior (e.g., Baker & Liu, 2021; Erdley & Asher, 1998; Gasser et al., 2012; Kochanska & Knaack, 2003). However, this is the first study from the social domain perspective to find that the maturity of children's moral concepts predicted responses to peer conflict during preschool. Gasser et al. found that aggressive and nonaggressive children did not differ on the criterion judgments. This may be because the children in this study were older than the current sample and could have had more mature moral concepts overall. It is also possible that differences in the criterion judgments were not found because the study did not differentiate between reactive aggressors and proactive aggressors.

Divergent results are observed when reactive and proactive aggressors are differentiated from each other. Whereas proactive aggressors showed marked deficits in their moral concepts, reactive aggressors had higher, not lower, moral reasoning (Jambon & Smetana, 2017). Yet these results are not incompatible with those of the current study. The current study measured

reasoning about retaliatory harm, whereas previous research measured reasoning about unprovoked harm. Jambon and Smetana suggest that reactively aggressive children may be more likely to condemn unprovoked harm due to their tendencies to attribute hostile intent (e.g., Crick & Dodge, 1996) and general experience engaging in retaliation against unprovoked harm. These children may subsequently see retaliation as a justified response; they view unprovoked harm as a serious offense, done with malintent, and as such they believe retaliation is a form of justice (Orbio de Castro et al., 2012). This reasoning is supported by research that has found that children who provide more justice-oriented justifications for moral transgressions are more likely to engage in reactive aggression (Baker & Liu, 2021).

However, moral reasoning only emerged as an independent predictor of competent responses at T1. T1 moral reasoning was correlaed with T2 competent responses, but it was not an independent longitudinal predictor of competent responses, nor was T2 moral reasoning correlated with competent responses at T2. The lack of a concurrent association at T2 is not surprising, given that participants' average criterion judgment ratings slightly decreased, while their competent responses increased. Yet is interesting to consider why moral reasoning did not improve over the course of the preschool year while both EF and SC did. In fact, preschool is a time marked by development across multiple different domains (Allen & Bickhard, 2018). So why did moral reasoning not follow this general trend? One possibility is that the "worse" criterion judgments at the end of preschool reflect a better understanding of the criterion questions and consequently, may be more indicative of their moral reasoning abilities. When confronted with provocation situations, young children tend to only focus on the outcome (Baird & Moses, 2001; Zelazo et al., 1996). Zelazo et al. suggest that this is because young children lack the EF ability to consider other information, such as cues regarding the transgressor's intentionality and other contextual information, in conjunction with the outcome of the situation. This inability to integrate all the pieces of relevant information results in immature reasoning about intentionality and emotional states (e.g., Caporaso et al., 2021; Cushman et al., 2013).

Moral reasoning requires similar integration of information, particularly in situations of retaliation (Baker et al., 2021; Richardson et al., 2012). Children not only need to consider the outcome of the actual transgression (i.e., the retaliatory act), but also the original unprovoked

transgression that prompted the retaliatory response. Criterion judgments then add an additional piece of information for children to consider: what if there was no rule against the action or if the teacher said it was ok? Young children may have a hard time thinking about all this information simultaneously to make an informed judgment. It seems unlikely that children are only focused on the outcome of the retaliatory act because then it would be expected that their acceptability judgments would be the same as their criterion judgments, a pattern that was not observed in the moral reasoning exploratory analyses. Rather, they may only focus on a few key pieces of the situation or confuse which act the contingency is referring to (e.g., believe the teacher is ok with the original transgression, not the retaliatory act). It is also possible the amount of information overwhelms some children so they answer at random, or based on a bias towards the positive end of the scale. As EF and other information processing abilities develop over the preschool year, children may be better at integrating the necessary pieces of information and make a judgment that might better reflect their true beliefs about retaliation. Perhaps the age-related trend in retaliation moral reasoning is non-linear. Young children seemingly have more advanced moral reasoning abilities, then pre-kindergarten children appear to have worse moral reasoning as cognitive skills develop. With further cognitive development, sophisticated forms of reasoning begin to emerge in middle or late childhood.

Indeed, T1 EF was related to T2 moral judgments, although in opposite directions depending on participants' level of classroom prosocial behavior. For participants with high levels of prosocial behavior, EF was negatively related to moral reasoning. Conversely, for participants with low levels of prosocial behavior, EF was positively related to moral reasoning. I speculate the high EF/high prosocial behavior group of participants are likely the most developmentally advanced participants. If lower moral reasoning at T2 is truly reflective of having a better understanding of the criterion questions, then it perhaps makes sense that this group of participants would have the lowest ratings. However, this interpretation begins to unravel when considering the high EF/low prosocial behavior group of participants. These participants still had high levels of EF so it stands to reason that they would perform the same as other high EF children if the decrease in moral reasoning was solely due to improvements in cognitive abilities. It also seems unlikely that

advanced reasoning ability alone would produce criterion judgments as low as the ones observed for the high prosocial behavior participants +1 standard deviation above the mean for EF.

Another interpretation of the data centers on the importance of social experience and lends support to the constructivist theory of moral development. According to the constructivist theory, children learn about moral rules through their social behavior and construct their moral concepts based on the messages they receive from their environment (Smetana et al., 2018). Based on this theory, children who are engaged in more problematic social encounters receive a higher amount of sociomoral feedback that can then be incorporated into their moral concepts (Dahl et al., 2018; Smetana et al., 2012). Many of the items on the prosocial subscale of the SCBE-30 inquire about children's abilities to cooperate with others, accept compromises, and use negotiation. Presumably, children who are rated low on these items are more likely to get into conflicts with their peers and consequently, receive more feedback about the appropriate way to respond to such conflicts.

Yet the process of incorporating this information into a mental concept likely requires a certain level of EF. EF may first allow children to attend to the moral messages in their environment (Smetana et al., 2012). Then, EF assists with assimilating that information into the moral concept through iterative reprocessing, the elaboration on and integration of information into existing mental concepts through deliberate reflection (Zelazo, 2015). Finally, EF supports the in-the-moment moral judgments when children are asked to reason about retaliation situations (Baker et al., 2021). Thus, mature reasoning about retaliation, at least in preschool, may depend on both EF and engagement in social conflict. Consequently, children with low EF and low prosocial behaviors have worse moral reasoning than children with similarly low prosocial behaviors and high EF because they lack the EF skills needed to use environmental feedback to inform their moral concepts and make complex moral judgments.

There are likely limits to the benefits of peer conflict situations for moral development. If children experience too much conflict and are unable to respond to these conflicts competently at least some of the time, this may result in lower moral reasoning. Indeed, Kochanska and colleagues have argued that early cognitive control abilities are important for moral development

because they provide children with opportunities to practice following moral rules (Kochanska & Knaack, 2003; Kochanska et al., 1997; Kochanska et al., 1996). This may be why classroom anger/aggression also negatively predicted moral reasoning. On the surface, these findings seemingly contradict the finding that prosocial behavior, as a main effect, is negatively related to moral reasoning. However, a conflict over a lack of cooperation may differ in both intensity and resolution compared to conflicts marked by aggression or anger. As such, children rated highly on the anger/aggression scale may have more persistent social difficulties in the classroom. They may also have lower EF, as indicated by the difficulty to control their emotions (e.g., Calkins & Marcovitch, 2010), so they may be the same children who are low EF/low prosocial behavior.

Nevertheless, this second interpretation is also inadequate when considering the high prosocial behavior participants. If it were simply the case that EF only mattered in the presence of mild conflict, it would be expected that the high and low EF participants with high levels of prosocial behavior would not differ in their moral reasoning. Yet the low EF/high prosocial participants had higher moral reasoning than the high EF/high prosocial behavior. It is unclear why the high EF/high prosocial behavior participants would have such low criterion judgments. Perhaps it's due to a combination of more advanced reasoning paired with a lack of social experience in conflict, but there may be other reasons for their low judgments.

Maybe these high EF/high prosocial children are particularly sensitive to the presence of rules and authority figures and are more likely to change their moral judgments to align with the rules or the teacher. Young children are typically able to understand that harm is wrong in contexts where no rule prohibits harm, or an authority figure explicitly says that the harm is permissible by the time they are 4 years of age (Smetana et al., 2014). However, this knowledge is typically assessed with unprovoked harm and not in retaliatory harm. In the presence of more complex retaliation judgments, high EF/high prosocial children may default to their knowledge about following rules and listening to authority figures to make their judgments. Regardless, the overall decrease in moral reasoning from T1 to T2 and a lack of an association between T2 moral reasoning and T2 competent behavior could simply be a reflection of the divergence between the two high EF groups over the course of the prekindergarten year.

Because EF is implicated in the development of moral reasoning, it is also possible that the longitudinal association between T1 criterion judgments and T2 competent responses was accounted for by T1 EF. Although criterion judgments and EF were only marginally correlated in the full T1 sample, they were significantly associated in the T1 longitudinal subsample. However, additional research is needed to test the possibility that EF mediates the longitudinal association between moral reasoning and SC.

The exploratory analyses further illustrated the importance of EF in moral reasoning. EF was particularly important for reasoning about relational and verbal retaliation, although a different pattern was found between high EF and low EF girls. It is possible that reasoning about relational and verbal retaliation requires more EF because these judgments are inherently more complex compared to physical aggression. Physical aggression is more salient during early childhood (Chen et al., 2001) and receives the most condemnation from teachers and parents (Bell & Willis, 2016; Waltzer et al., 2019). Moral messages about verbal and relational aggression may also be more subtle or inconsistent so greater EF ability may be required to pick up on these messages in the environment. In addition, preschool children generally have less experience with relational aggression, as frequency of relational aggression peaks in late childhood (Orpinas et al., 2015; Vaillancourt et al., 2007), and that could make it harder to reason about. However, high EF girls had the lowest level of moral reasoning for physical aggression. Perhaps this is another example of the importance of social experience in moral development. High EF girls likely do not engage in physical aggression themselves, at least not to the same degree as low EF girls and boys (e.g., Caporaso et al., 2019; Poland et al., 2016), so it may be the case that they have less mature moral concepts regarding physical aggression.

The exploratory analyses also revealed that preschool children's moral reasoning about specific types of retaliation only related to some forms of aggression, both concurrently and longitudinally. Children's judgments regarding physical retaliation (i.e., hit) were negatively correlated with hit responses to conflict, but judgments related to verbal (i.e., yell) and relational retaliation did not correlate with their respective aggressive responses on the VSG. Interestingly, the hit response option was also negatively correlated with T1 relational retaliation judgments (at T1 and T2) and T1 verbal judgments (at T2). It is unclear why judgments specifically regarding

verbal aggression and relational aggression would relate to the hit response option but not their respective response options. Perhaps participants with higher criterion judgments across all retaliation types had high enough moral reasoning to know to avoid choosing the hit response, because of its salient immoral status (Murray-Close et al., 2006; Smetana et al., 2003; Smetana et al., 1999), but not high enough to avoid the other aggressive responses. Said another way, participants who chose the hit response may have globally lower moral reasoning. But because the relational and yell response options are less morally salient, moral reasoning does not differentiate between children who choose or do not choose those responses.

In general, the exploratory analyses revealed that children do not have mature moral concepts regarding retaliation during preschool. This is illustrated by the significant difference between acceptability judgments and the two criterion judgments. The maturity of the moral concepts also did not differ by response type. It was hypothesized that children may have a more advanced moral concept for physical retaliation compared to psychological retaliation, but the lack of a judgment x retaliation type interaction suggests that physical retaliation followed the same overall pattern captured by the judgment type main effect. This finding is consistent with previous research that has found that moral concepts for retaliation do not reach maturity for various forms of aggression until late childhood (Smetana et al., 2003). This protracted trajectory compared to reasoning about unprovoked harm may be due to greater EF demand placed on retaliation judgments, but additional research is required to examine the specific factors that lead to these different developmental trajectories.

Limitations and Future Directions

Further research is needed to examine the findings of the current study due to several limitations. Notably, the longitudinal aspect of the study was disrupted by COVID-19 and many of the T1 participants were unable to be tested at T2. Despite considerable effort to retain as many participants as possible, the T2 sample is much smaller than originally planned. The results observed in the current study should therefore be interpreted with caution.

Relatedly, another limitation of the current study is the use of measures that differed from those that were originally intended between T1 and T2, and the difference in administration methods between the two time points. Because testing had to be entirely virtual for T2, some measures were either adapted so they could be presented virtually (i.e., MST), or measures were switched out entirely (i.e., DCCS). Children could have found these measures harder (or easier) to complete in the virtual format than the versions they completed at T1. Measures that did not change presentation mode (e.g., the VSG and the social events interview were presented on Qualtrics at both time points) may have been elicited different types of responses in the virtual format. The two social tasks, in particular, could have elicited more demand characteristics from children. During virtual testing sessions, parents were instructed to sit near their child so they could assist with any technology-related issues. The presence of their parents could have been a salient reminder to some participants that they should choose more "acceptable" responses on the social measures. However, this seems somewhat unlikely, at least for the social events interview, given the slight decrease in criterion judgments between time points.

The use of virtual provocation situations also limits the conclusions that can be drawn from this study. Although there are associations between similar measures and EF (Caporaso et al., 2019; Denham et al., 2014), emotional responses to provocation (Denham et al., 2013), and developmental outcomes such as school readiness (Denham & Bassett, 2018), there is still an issue of external validity. Notably, none of the classroom behavior measures related to participants' performance on the VSG. It is possible that there could be an issue with the classroom behavior measures chosen for the current study, as they also did not correlate with many of the other variables in the study. For example, the response to provocation and reactive aggression subscales did not correlate with T2 EF even though similar questionnaires used to assess children's aggression typically finds this correlation (e.g., Poland et al., 2016; Rathert et al., 2011). Nonetheless, additional research is needed to determine if the VSG truly captures children's real-life social problem skills or if it only measures children's knowledge of what constitutes an appropriate response to conflict.

The results from the current study can inspire many additional avenues of future research. First, additional longitudinal research that covers a wider developmental period could be used to

investigate the emergence and expression accounts of the EF and SC association further, as well as potential moderators of this association. Perhaps EF assists with the emergence of SC for most children at the beginning of the observed decrease in aggression (around 3.5 years of age; Alink et al., 2006; Carbonneau et al., 2015), but once children are closer to 5 years of age, the emergence mechanism is only important for children high in surgency. However, the current study did not identify any factors indicative of the group of children who are never or rarely aggressive. Future research should include additional factors as potential moderators, such as parental emotion socialization and the presence of older siblings in the home, to explore the possibility that there is indeed a group of young children who do require EF to be competent. Temperament measures that break down the three dimensions to their subcomponents may also be useful in examining the relative contributions of temperament to the development of SC and how these factors may affect the EF-SC relation over time.

Another potential line of research would be to explore how and why the components of EF differentially relate to SC development during early childhood. The current study found preliminary evidence that response inhibition was a more important EF component for early SC but did not predict or relate to later SC. Perhaps these results are due to changes in the structure of EF during early childhood (Lehto et al., 2003). Perhaps these results indicate that children use EF differently as they develop SC. While a provocative finding, much more research is needed, with multiple measures of each EF component, to understand why this finding emerged in the current study.

Finally, a third line of future research would examine the development of children's moral reasoning about retaliation. Additional research is needed to understand how this moral reasoning develops over the course of childhood, how EF and social experiences relate to this development, and whether there is a nonlinear trend due to the development of moral reasoning about retaliation. The role of children's everyday experiences in moral development and how EF interacts with these experiences can also be further explored. An observation study in preschool classrooms that could examine the types of interactions children have with peers, as well as the feedback they receive from their teachers, could answer some of the remaining questions about the differential results observed based on retaliation type. Indeed, when asked to rank each

response type to provocation from best to worst, some teachers commented that they were unsure how to classify the "yell" response because on one hand, they encourage children to use their words over physical aggression, but on the other hand, they place an emphasis on nice words. Some also reported that it is dependent on the particular child and their current level of SC. An examination of the types of messages children receive from their teachers, as well as their parents, may provide further insight into individual differences in children's moral reasoning about retaliation.

Conclusion

The current study provided an in-depth examination of the development of SC and moral reasoning during the final year of preschool. Children's EF is important for the emergence of SC in conflict situations, but particularly for children who may struggle with impulsive social behavior. Otherwise, children may only need EF in the moment of conflict to recognize the most optimal response options and possibly to enact those responses. Additionally, the results of the current study demonstrate the importance of children's social experiences and EF in the development of moral reasoning, and that moral reasoning could contribute to their ability to respond to provocation appropriately. These results inspire lines of future research, all with the goal of understanding the factors that promote optimal social and moral development across early childhood.

REFERENCES

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions.* Sage Publications, Inc.

Alduncin, N., Huffman, L. C., Feldman, H. M., & Loe, I. M. (2014). Executive function is associated with social competence in preschool-aged children born preterm or full term. *Early Human Development, 90*(6), 299–306. doi: 10.1016/j.earlhumdev.2014.02.011

Alink, L.R. A., Mesman, J., Zeijl, J. V., Stolk, M. N., Juffer, F., & Koot, H. M. (2006). The early childhood aggression curve: Development of physical aggression in 10- to 50-month-old children. *Child Development, 77*(4), 954-966. doi: 10.1111/j.1467-8624.2006.00912.x

Allen, J. W. P. & Bickhard, M. (2018). Stage fright: Internal reflection as a domain general enabling constraint on the emergence of explicit thought. *Cognitive Development, 45*, 77-91. doi: 10.1016/j.cogdev.2017.12.005

Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and visuospatial short-term and working memory in children: Are they separable? *Child Development, 77*, 1698–1716. doi: 10.1111/j.1467-8624.2006.00968.x

Arsenio, W. F., Adams, E., & Gold, J. (2009). Social information processing, moral reasoning, and emotion attribution: Relations with adolescents' reactive and proactive aggression. *Child Development, 80*, 1739-1755. doi: 10.1111/j.1467-8624.2009.01365.x

Arsenio, W. F., & Lemerise, E. A. (2004). Aggression and moral development: Integrating social information processing and moral domain models. *Child Development, 75*(4), 987–1002. doi: 10.1111/j.1467-8624.2004.00720.x

Baird, J. A., & Moses, L. J. (2001). Do preschoolers appreciate that identical actions may be motivated by different intentions? *Journal of Cognition & Development, 2*, 413–448. doi: 10.1207/S15327647JCD02044

Baker, E. R., D'Esterre, A. P., & Weaver, J. P. (2021). Executive function and theory of mind in explaining young children's moral reasoning: A test of the hierarchical competing systems model. *Cognitive Development (58),* 1-16. doi: 10.1016/j.cogdev.2021.101035

Baker, E. R. & Liu, Q. (2021). Moral reasoning and moral behavior: Intersections of reasoning with aggressive forms and functions in early childhood. *Early Education and Development, 32*, 534-552. doi: 10.1080/10409289.2020.1780561

Beauchamp, M. H., & Anderson, V. (2010). SOCIAL: An integrative framework for the development of social skills. *Psychological Bulletin, 136,* 39–64. doi: 10.1037/a0017768

Bell, K. J. S. & Willis, W. G. (2016). Teachers' perceptions of bullying among youth. *The Journal of Education Research, 2,* 159-168. doi: 10.1080/00220671.2014.931833

Berdan, L. E., Keane, S. P., & Calkins, S. D. (2008). Temperament and externalizing behavior: Social preference and perceived acceptance as protective factors. *Developmental Psychology, 44,* 957–968. doi: 10.1037/0012-1649.44.4.957

Blair, C. (2014). Stress and the development of executive functions: Experiential canalization of brain and behavior. In P. Zelazo and M. Sera (Eds.), *37th Minnesota symposium on child psychology: Developing cognitive control processes: Mechanisms, implications, and interventions* (pp. 145-180). Hoboken NJ: Wiley.

Blankemeyer, M., Culp, R. E., Hubbs-Tait, L., & Culp, A. M. (2002). A tool for identifying preschoolers' deficits in social competence: The Preschool Taxonomy of Problem Situations. *Education and Treatment of Children, 25,* 208-223.

Bowlby, J. (1982). *Attachment and loss: Vol. 1. Attachment* (2nd ed.). New York: Basic Books.

Brook, J. S. & Newcomb, M. D. (1995). Childhood aggression and unconventionality: Impact on later academic achievement, drug use, and workforce involvement. *The Journal of Genetic Psychology, 156*(4), 393-410.

Burks, V. S., Laird, R. D., Dodge, K. A., Pettit, G. S., & Bates, J. E. (1999). Knowledge structures, social information processing, and children's aggressive behavior. *Social Development, 8*(2), 220–236. doi: 10.1111/1467-9507.00092

Calkins, S. D., & Marcovitch, S. (2010). Emotion regulation and executive functioning in early development: Integrated mechanisms of control supporting adaptive functioning. In S. D. Calkins & M. A. Bell (Eds.), *Human brain development. Child development at the intersection of emotion and cognition* (pp. 37–57). Washington, DC: American Psychological Association.

Capaldi, D. M. (1992). Co-occurrence of conduct problems and depressive symptoms in early adolescent boys: II. A 2-year follow-up at Grade 8. *Development and Psychopathology, 4,* 125–144. doi: 10.1017/S0954579400005605

Caporaso, J. S., Boseovski, J. J., & Marcovitch, S. (2019). The individual contributions of three executive functions to preschool social competence. *Infant and Child Development*, 1-17. doi: 10.1002/icd.2132

Caporaso, J. S. & Marcovitch, S. (2017, October). *Moral understanding and executive function in the resolution of preschool peer conflict.* Poster presented at the biennial meeting of the Cognitive Development Society, Columbus, OH.

Caporaso, J. S. & Marcovitch, S. (2021). The effect of taxing situations on preschool children's responses to peer conflict. *Cognitive Development, 57*, 1-14. doi: 10.1016/j.cogdev.2020.100989

Caporaso, J. S., Marcovitch, S., & Boseovski, J. J. (2021). Executive function and the development of social information processing during preschool. *Cognitive Development, 58*, 1-15. doi: 10.1016/j.cogdev.2021.101018

Carbonneau, R., Boivin, M., Brendgen, M., Nagin, D., & Tremblay, R. E. (2016). Comorbid development of disruptive behaviors from age 1½ to 5 years in a population birth-cohort and association with school adjustment in first grade. *Journal of Abnormal Child Psychology, 44*, 677–690. doi: 10.1007/s10802-015-0072-1

Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology, 28*, 595-616. doi: 10.1207/s15326942dn2802_3

Carlson, S. M., Claxton, L. J., & Moses, L. J. (2015). The relation between executive function and theory of mind is more than skin deep. *Journal of Cognition and Development, 16,* 186-197. doi: 10.1080/15248372.2013.824883

Carlson, S. M., Mandell, D. J., & Williams, L. (2004). Executive function and theory of mind: Stability and prediction from ages 2 to 3. *Developmental Psychology, 40*, 1105–1122. doi: 10.1037/0012-1649.40.6.1105

Chen, D. W., Fein, G. G., Killen, M., & Tam, H. (2001). Peer conflicts of preschool children: Issues, resolution, incidence, and age-related patterns, *Early Education and Development, 12*, 523-544. doi: 10.1207/s15566935eed1204

Ciairano, S., Bonino, S., & Miceli, R. (2006). Cognitive flexibility and social competence from childhood to early adolescence. *Cognition, Brain, Behavior 10*, 343–366.

Cohen, J. S. & Mendez, J. L. (2009). Emotion regulation, language ability, and the stability of preschool children's peer play behavior. *Early Education and Development, 20*, 1016-1037. doi: 10.1080/10409280903305716

Cole, P. M., Tan, P. Z., Hall, S. E., Zhang, Y., Crnic, K. A., Blair, C. B., & Li, R. (2011). Developmental changes in anger expression and attention focus: learning to wait. *Developmental Psychology, 47*, 1078–1089. doi: 10.1037/a0023813

Cole, P. M., Usher, B. A., & Cargo, A. P. (1993) Cognitive risk and its association with risk for disruptive behavior disorder in preschoolers. *Journal of Clinical Child Psychology, 22*, 154-164. doi: 10.1207/ s15374424jccp2202_3

Coolahan, K., Fantuzzo, J., Mendez, J. L., & McDermott, P. (2000). Preschool peer interactions and readiness to learn: Relationships between classroom peer play and learning behaviors and conduct. *Journal of Educational Psychology, 92*, 458–465. doi: 10.1037/0022-0663.92.3.458

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

Crick, N. R. & Dodge, K. A. (1996). Social information-processing mechanisms in reactive and proactive aggression. *Child Development, 67*, 993-1002. doi: 10.1111/j.1467-8624.1996.tb01778.x

Crick, N. R., Ostrov, J. M., Burr, J. E., Cullerton-Sen, C., Jansen-Yeh, E. & Ralston, P. (2006). A longitudinal study of relational and physical aggression in preschool. *Journal of Applied Developmental Psychology, 27*, 254-268. doi: 10.1016/j.appdev.2006.02.006

Crick, N. R., & Werner, N. E. (1998). Response decision processes in relational and overt aggression. *Child Development, 69*, 1630–1639. doi: 10.2307/1132136

Cushman, F., Sheketoff, R., Wharton, S., & Carey, S. (2013). The development of intent-based moral judgement. *Cognition, 127*, 6-21. doi: 10.1016/j.cognition.2012.11.008

Dahl, A., Waltzer, T., & Gross, R. L. (2018). Helping, hitting, and developing: Toward a constructivist- interactionist account of early morality. In C.C. Helwig (Ed.), *New perspectives on moral development* (pp. 33–53). New York, NY: Routledge.

Degnan, K. A., Hane, A. A., Henderson, H. A., Moas, O. L., Reeb-Sutherland, B. C., & Fox, N. A. (2011). Longitudinal stability of temperamental exuberance and social–emotional outcomes in early childhood. *Developmental Psychology, 47*, 765–780. doi: 10.1037/a0021316

Denham, S. A., & Bassett, H. H. (2018). 'You hit me! That's not nice and it makes me sad!!': Relations of young children's social information processing and early school success. *Early Child Development and Care*, 1-15. doi: 10.1080/03004430.2018.1491562

Denham, S. A., Bassett, H. H., Way, E., Kalb, S., Warren-Khot, H., & Zinsser, K. (2014). "How would you feel? What would you do?" Development and underpinnings of preschoolers' social information processing. *Journal of Research in Childhood Education, 28*, 182–202. doi: 10.1080/02568543.2014.883558

Denham, S. A., Bassett, H. H., & Wyatt, T. M. (2010). Gender differences in the socialization of preschoolers' emotional competence. *New Directions for Child and Adolescent Development, 2010,* 29–49. doi: 10.1002/cd.267

Denham, S.A., Bouril, B., & Belouad, F. (1994). Preschoolers' affect and cognition about challenging peer situations. *Child Study Journal, 24*, 1-22.

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.x

DeWall, N. C., Finkel, E. J., & Denson, T. F. (2011). Self-control inhibits aggression. *Social and Personality Psychology Compass 5*, 458–472. doi: 10.1111/j.1751-9004.2011.00363.x

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64,* 135-168. doi: 10.1146/annurev-psych-113011-143750

Dodge, K. A., Pettit, G.S., McClaskey, C.L., Brown, M.M., & Gottman, J.M. (1986). Social competence in children. *Monographs of the Society for Research in Child Development, 51*, I-85. doi:10.2307/1165906

Dodge, K. A., & Rabiner, D. L. (2004). Returning to roots: On social information processing and moral development. *Child Development, 75*, 1003-1008. doi: 10.1111/j.1467-8624.2004.00721.x

Dollar J. M. & Calkins S. D. (2019). The development of anger. In: LoBue V., Pérez-Edgar K., Buss K. (eds) *Handbook of Emotional Development*. Springer, Cham. doi: 10.1007/978-3-030-17332-69

Dollar, J. M. & Stifter, C. A. (2012). Temperamental surgency and emotion regulation as predictors of childhood social competence. *Journal of Experimental Child Psychology, 112*, 178–194. doi: 10.1016/j.jecp.2012.02.004

Domitrovich, C. E., Cortes, R. C., & Greenberg, M. T., (2007). Improving young children's social and emotional competence: A randomised trial of the preschool "PATHS" curriculum. *The Journal of Primary Prevention, 28*,67–91. doi: 10.1007/s10935-007-0081-0

Eisenberg, N., Cumberland, A., Spinrad, T. L., Fabes, R. A., Shepard, S. A., Reiser, M., Murphy, B. C., Losoya, S. H., Guthrie, I. K. (2001) The relations of regulation and emotionality to children's externalizing and internalizing problem behavior. *Child Development, 72*, 1112-1134. doi: 10.1111/1467-8624.00337.

Eisenberg, N., Fabes, R. A., & Murphy, B. C. (1996). Parents' reactions to children's negative emotions: relations to children's social competence and comforting behavior. *Child Development, 67*, 2227–2247. doi: 10.1111/j.1467-8624.1996.tb01854.x

Eisenberg, N., Fabes, R. A., Murphy, B., Maszk, P., Smith, M., & Karbon, M. (1995). The role of emotionality and regulation in children's social functioning: A longitudinal study. *Child Development, 66*, 1360–1384. doi: 10.2307/1131652

Eisenberg, N., Fabes, R. A., Nyman, M., Bernzweig, J., & Pinuelas, A. (1994). The relations of emotionality and regulation to children's anger-related reactions. *Child Development, 65*, 109–128. doi: 10.2307/1131369

Eisenberg, N., Sadovsky, A., Spinrad, T. L., Fabes, R. A., Losoya, S. H., Valiente, C., Reiser, M., Cumberland, A., & Shepard, S. A. (2005). The relations of problem behavior status to children's negative emotionality, effortful control, and impulsivity: Concurrent relations and prediction of change. *Developmental Psychology, 41*, 193–211. doi: 10.1037/0012-1649.41.1.193

Eisenberg, N., Valiente, C., Morris, A. S., Fabes, R. A., Cumberland, A., Reiser, M., Gershoff, E. T., Shepard, S. A., & Losoya, S. (2003). Longitudinal relations among parental emotional expressivity, children's regulation, and quality of socioemotional functioning. *Developmental Psychology, 39*, 3–19. doi: 10.1037/0012-1649.39.1.3

Eisenberg, N., Valiente, C., Spinrad, T. L., Cumberland, A., Liew, J., Reiser, M., Zhou, Q., & Losoya, S. H. (2009). Longitudinal relations of children's effortful control, impulsivity, and negative emotionality to their externalizing, internalizing, and co-occurring behavior problems. *Developmental Psychology, 45*, 988–1008. doi: 10.1037/a0016213

Ellis, M. L., Weiss, B., & Lochman, J. E. (2009). Executive functions in children: Associations with aggressive behavior and appraisal processing. *Journal of Abnormal Child Psychology, 37*, 945–956. doi: 10.1007/s10802-009-9321-5

Erdley, C. A., & Asher, S. R. (1996). Children's social goals and self-efficacy perceptions as influences on their responses to ambiguous provocation. *Child Development, 67*, 1329–1344.

Erdley, C. A., & Asher, S. R. (1998). Linkages between children's beliefs about the legitimacy of aggression and their behavior. *Social Development, 7*, 321-339. doi: 10.1111/1467-9507.00070

Fabes, R. A., Leonard, S. A., Kupanoff, K., & Martin, C. L. (2001). Parental coping with children's negative emotions: relations with children's emotional and social responding. *Child Development, 72*, 907–920. doi: 10.1111/1467-8624.00323

Fontaine, R. G & Dodge, K.A. (2006). Real-time decision making and aggressive behavior in youth: A heuristic model of response evaluation and decision (RED). *Aggressive Behavior, 32*, 604-624. doi: 10.1002/ab

Gagne, J. R., Miller, M. M., & Goldsmith, H. H. (2013). Early-but modest-gender differences in focal aspects of childhood temperament. *Personality and Individual Differences, 55*, 95–100. doi: 10.1016/j.paid.2013.02.006

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: a review using an integrative framework. *Psychological Bulletin, 134*, 31–60. doi: 10.1037/0033-2909.134.1.31

Gasser, L., Malti, T., & Gutzwiller-Helfenfinger, E. (2012). Aggressive and nonaggressive children's moral judgments and moral emotion attributions in situations involving retaliation and unprovoked aggression. *The Journal of Genetic Psychology: Research and Theory on Human Development, 173*, 417–439. doi: 10.1080/00221325.2011.614650

Gilliom, M., Shaw, D. S., Beck, J. E., Schonberg, M. A., & Lukon, J. L. (2002). Anger regulation in disadvantaged preschool boys: Strategies, antecedents, and the development of self-control. *Developmental Psychology, 38*, 222–235. doi: 10.1037/0012-1649.38.2.222

Granvald, V., & Marciszko, C. (2016). Relations between key executive functions and aggression in childhood. *Child Neuropsychology, 25*, 537-555. doi: 10.1080/09297049.2015.1018152

Gunnar, M. R., Sebanc, A. M., Tout, K., Donzella, B., & van Dulmen, M. M. (2003). Peer rejection, temperament, and cortisol activity in preschoolers. *Developmental Psychobiology, 43*, 346–358. doi: 10.1002/dev.10144

Hala, S., Pexman, P., Climie, E., Rostad, K., & Glenwright, M. (2010). A bidirectional view of executive function and social interaction. In B. W. Sokol, U. Müller, J. I. M. Carpendale, A. R. Young, & G. Iarocci (Eds), *Self and social regulation: Social interaction and the development of social understanding and executive functions* (pp. 293-311). New York: Oxford University Press

Hill, A. L., Degnan, K. A Calkins, S. D., & Keane, S. P. (2006). Profiles of externalizing behavior problems for boys and girls across preschool: The roles of emotion regulation and inattention. *Developmental Psychology, 42*, 913–28. doi: 10.1037/0012-1649.42.5.913

Hughes, C., White, A., Sharpen, J., & Dunn, J. (2000). Antisocial, angry, and unsympathetic: "Hard-to-manage" preschoolers' peer problems and possible cognitive influences. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 41*, 169–179. doi: 10.1111/1469-7610.00558

Jambon, M. & Smetana, J. G. (2017). Individual differences in prototypical moral and conventional judgements and children's proactive and reactive aggression. *Child Development,* 1-17. doi: 10.1111/cdev.12757

Jones, S., Eisenberg, N., Fabes, R. A., & MacKinnon, D. P. (2002). Parents' reactions to elementary school children's negative emotions: Relations to social and emotional functioning at school. *Merrill-Palmer Quarterly, 48*, 133–159. doi: 10.1353/mpq.2002.0007

Kochanska, G. & Aksan, N. (2006). Children's conscience and self-regulation. *Journal of Personality, 74*, 1587-1617. doi: 10.1111/j.1467-6494.2006.00421.x

Kochanska, G. & Knaack, A. (2003). Effortful control as a personality characteristic of young children: Antecedents, correlates, and consequences. *Journal of Personality, 71*, 1087–1112. doi: 10.1111/1467-6494.7106008

Kochanska, G., Murray, K., & Coy, K. C. (1997). Inhibitory control as a contributor to conscience in childhood: From toddler to early school age. *Child Development, 68*, 263-277. doi: 10.1111/j.1467-8624.1997.tb01939.x

Kochanska, G., Murray, K., Jacques, T. Y., Koenig, A. L., & Vandegeest, K. A. (1996). Inhibitory control in young children and its role in emerging internalization. *Child Development, 67*, 490–507. doi: 10.2307/1131828

LaFreniere, P. J., & Dumas, J. E. (1996). Social competence and behavior evaluation in children ages 3 to 6 years: The short form (SCBE-30). *Psychological Assessment, 8*, 369-377. doi: 10.1037/1040-3590.8.4.369

Lagattuta, K. H., Sayfan, L., & Monsour, M. (2011). A new measure for assessing executive function across a wide age range: Children and adults find happy-sad more difficult than day-night. *Developmental Science, 14*, 481–489. doi: 10.1111/j.1467-7687.2010.00994.x

Landry, S. H., Smith, K. E., & Swank, P. R. (2009). New directions in evaluating social problem solving in childhood: Early precursors and links to adolescent social competence. In C. Lewis & J. I. M. Carpendale (Eds.), Social interaction and the development of executive function. *New Directions in Child and Adolescent Development, 123*, 51–68. doi: 10.1002/cd

Landry, S. H., & Smith, K. E. (2010). Early social and cognitive precursors and parental support for self-regulation and executive function: relations from early childhood into adolescence. In B. W. Sokol, U. Müller, J. I. M. Carpendale, A. R. Young, & G. Iarocci (Eds), *Self and social regulation: Social interaction and the development of social understanding and executive functions* (pp. 386-417). New York: Oxford University Press

Lansford, J. E., Malone, P. S., Dodge, K. A., Pettit, G. S., & Bates, J. E. (2010). Developmental cascades of peer rejection, social information processing biases, and aggression during middle childhood. *Developmental Psychopathology, 22*, 593-602. doi: 10.1017/S0954579410000301

Lehto, J. E., Juujaervi, P., Kooistra, L., & Pulkkinen, L. (2003). Dimensions of executive functioning: Evidence from children. *British Journal of Developmental Psychology, 21,* 59–80. doi: 10.1348/026151003321164627

Lemerise, E., & Arsenio, W. (2000). An integrated model of emotion processes and cognition in social information processing. *Child Development, 71*, 107–1118. doi: 10.1111/1467-8624.00124

Marcovitch, S., O'Brien, M., Calkins, S. D., Leerkes, E. M., Weaver, J. M., & Levine, D. W. (2015). A longitudinal assessment of the relation between executive function and theory of mind at 3, 4, and 5 years. *Cognitive Development, 33,* 40-55. doi: 10.1016/j.cogdev.2014.07.001

McQuade, J. D., Murray-Close, D., Shoulberg, E. K., & Hoza, B. (2013). Working memory and social functioning in children. *Journal of Experimental Child Psychology, 115*, 422–435. doi: 10.1016/j.jecp.2013.03.002

Mendez, J. L., Fantuzzo, J., & Cicchetti, D. (2002). Profiles of social competence among low-income African American preschool children. *Child Development, 73,* 1085-1100. doi: 10.1111/1467-8624.00459

Myers, T. A. (2011). Goodbye, listwise deletion: Presenting hot deck imputation as an easy and effective tool for handling missing data. *Communication Methods and Measures, 5*, 297-310. doi: 10.1080/19312458.2011.624490

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: a latent variable analysis. *Cognitive Psychology, 41*, 49–100. doi: 10.1006/cogp.1999.0734

Moran, L. R., Lengua, L. J., & Zalewski, M. (2013). The interaction between negative emotionality and effortful control in early social-emotional development. *Social Development, 22*, 340-362. doi: 10.1111/sode.12025.

Moffitt, T.E., Caspi, A., Rutter, M., & Silva, P. (2001). *Sex differences in antisocial behavior: Conduct disorder, delinquency and violence in the Dunedin Longitudinal Study*. Cambridge: Cambridge University Press

Moses, L. J. (2001). Executive accounts of theory-of-mind development. Commentary on "Meta-analysis of theory-of-mind development: The truth about false belief." *Child Development, 72*, 688–690. doi: 10.1111/1467-8624.00306

Müller, U., Liebermann-Finestone, D. P., Carpendale, J. I., Hammond, S. I., Bibok, M. B. (2012). Knowing minds, controlling actions: the developmental relations between theory of mind and executive function from 2 to 4 years of age. *Journal of Experimental Child Psychology, 111*, 331-348. doi: 10.1016/j.jecp.2011.08.014

Murray-Close, D., Crick, N. R., & Galotti, K. M. (2006). Children's moral reasoning regarding physical and relational aggression. *Social Development, 15*, 345–372. doi: 10.1111/j.1467-9507.2006.00346.x

Murray-Close, D., Crick, N. R., Tseng, W.-L., Lafko, N., Burrows, C., Pitula, C., & Ralston, P. (2014). Physiological stress reactivity and physical and relational aggression: The moderating roles of victimization, type of stressor, and child gender. *Development and Psychopathology, 26*, 589–603. doi: 10.1017/S095457941400025X

NICHD Early Child Care Research Network. (2004). Trajectories of Physical Aggression from Toddlerhood to Middle Childhood: Predictors, Correlates, and Outcomes. *Monographs of the Society for Research in Child Development, 69*, I-143. Retrieved from http://www.jstor.org/stable/3701390

Nigg, J. T., Quamma, J. P., Greenberg, M. T., & Kusche, C. A. (1999). A two-year longitudinal study of neuropsychological and cognitive performance in relation to behavioral problems and competencies in elementary school children. *Journal of Abnormal Child Psychology, 27*, 51–63. doi: 10.1023/A:1022614407893

O'Toole, S. E., Tsermentseli, S., Humayun, S., & Monks, C. P. (2019). Cool and hot executive functions at 5 years old as predictors of physical and relational aggression between 5 and 6 years old. *International Journal of Behavioral Development, 43*, 157–165. doi: 10.1177/0165025418798498

Olson, S. L., Lopez-Duran, N., Lunkenheimer, E. S., Chang, H., & Sameroff, A. J. (2011). Individual differences in the development of early peer aggression: Integrating contributions of self-regulation, theory of mind, and parenting. *Development and Psychopathology, 23,* 253-266. doi:10.1017/S0954579410000775

Olson, S. L., Sameroff, A. J., Kerr, D. C., Lopez, N. L., Wellman, H. M. (2005). Developmental foundations of externalizing problems in young children: The role of effortful control. *Developmental Psychopathology, 17,* 25-45. doi: 10.1017/s0954579405050029

Orobio de Castro, B. (2004). The development of social information processing and aggressive behaviour: Current issues. *European Journal of Developmental Psychology, 1*, 87–102. doi: 10.1080/17405620444000058

Orobio de Castro, B., Verhulp, E. E., & Runions, K. (2012). Rage and revenge: Highly aggressive boys' explanations for their responses to ambiguous provocation. European *Journal of Developmental Psychology, 9*, 331–350. doi: 10.1080/17405629.2012.680304

Orpinas, P., McNicholas, C., & Nahapetyan, L. (2015). Gender differences in trajectories of relational aggression perpetration and victimization from middle to high school. *Aggressive Behavior, 41*, 401–412. doi: 10.1002/ab.21563

Ostrov, J. M., Kamper, K. E., Hart, E. J., Godleski, S. A., & Blakely-McClure, S. J. (2014). A gender-balanced approach to the study of peer victimization and aggression subtypes in early childhood. *Developmental Psychopathology, 26*, 575-587. doi: 10.1017/S0954579414000248

Pahl, K., & Barrett, P. (2010). Preventing anxiety and promoting social and emotional strength in preschool children: A universal evaluation of the Fun FRIENDS program. *Advances in School Mental Health Promotion, 3,* 14-25. doi: 10.1080/1754730X.2010.9715683

Premo, J. E., & Kiel, E. J. (2014). The effect of toddler emotion regulation on maternal emotion socialization: Moderation by toddler gender. *Emotion, 14*, 782–793. doi: 10.1037/a0036684

Pinderhouse, E. E., Dodge, K. Zelli, A. Bates, J. E., Pettit, G. S. (2000) Discipline responses: Influence of parents' socioeconomic status, ethnicity, beliefs about parenting, stress and cognitive-emotional processes, *Journal of Family Psychology, 14*, 380-400. doi: 10.1037//0893-3200.14.3.380

Poland, S. E., Monks, C. P., & Tsermentseli, S. (2016). Cool and hot executive function as predictors of aggression in early childhood: Differentiating between the function and form of aggression, *British Journal of Developmental Psychology, 34*, 181–197. doi: 10.1111/bjdp.12122

Putnam, S. P., & Rothbart, M. K. (2006). Development of short and very short forms of the children's behavior questionnaire. *Journal of Personality Assessment, 87*, 102-112. doi: 10.1207/s15327752jpa8701_09

Raikes, H. A., Virmani, E. A, Thompson, R. A., & Hatton, H. (2013). Declines in peer conflict from preschool through first grade: Influences from early attachment and social information processing. *Attachment and Human Development, 15*, 65–82. doi: 10.1080/14616734.2012.728381

Razza, R. A., & Blair, C. B. (2009). Associations among false-belief understanding, executive function, and social competence: A longitudinal analysis. *Journal of Applied Developmental Psychology, 30*, 332–343. doi: 10.1016/j.appdev.2008.12.020

Rathert, J., Fite, P. J., & Gaertner, A. E. (2011). Associations between effortful control, psychological control and proactive and reactive aggression. *Child Psychiatry and Human Development, 42*, 609–621. doi: 10.1007/s10578-011-0236-3

Richardson, C. B., Mulvey, K. L., & Killen, M. (2012). Extending social domain theory with a process-based account of moral judgments. *Human Development, 55*, 4–25. doi: 10.1159/000335362

Riggs, N. R., Jahromi, L. B., Razza, R. P., Dillworth-Bart, J. E., & Müller, U. (2006). Executive function and the promotion of social-emotional competence. *Journal of Applied Developmental Psychology, 27*, 300–309. doi: 10.1016/j.appdev.2006.04.002

Rhoades, B. L., Greenberg, M. T., & Domitrovich, C. E. (2009). The contribution of inhibitory control to preschoolers' social-emotional competence. *Journal of Applied Developmental Psychology, 30*, 310–320. doi: 10.1016/j.appdev.2008.12.012

Roman, A. S., Pisoni, D. B., & Kronenberger, W. G. (2014). Assessment of working memory capacity in preschool children using the missing scan task. *Infant and Child Development, 23*, 575-583. doi: 10.1002/icd.1849

Rothbart, M. K., Ahadi, S. A., & Evans, D. E. (2000). Temperament and personality: Origins and outcomes. *Journal of Personality and Social Psychology, 78*, 122-35. doi: 10.1037//0022-3514.78.1.122

Rothbart, M. K., & Bates, J. E. (2006). Temperament. In N. Eisenberg, W. Damon, & R. M. Lerner (Eds.), *Handbook of child psychology: Social, emotional, and personality development* (pp. 99–166). Hoboken NJ: Wiley.

Rubin, K., & Rose-Krasnor, L. (1992). Interpersonal problem solving and social competence in children. In V. B. van Hasselt & M. Hersen (Eds.), *Handbook of social development: A lifespan perspective*. New York: Plenum.

Runions, K. C., & Keating, D. P. (2010). Anger and inhibitory control as moderators of children's hostile attributions and aggression. *Journal of Applied Developmental Psychology, 31,* 370–378. doi: 10.1016/j.appdev.2010.05.006

Schoemaker, K., Bunte, T., Espy, K. A., Deković, M., & Matthys, W. (2014). Executive functions in preschool children with ADHD and DBD: An 18-month longitudinal study. *Developmental Neuropsychology, 39,* 302–15. doi: 10.1080/87565641.2014.911875

Sebanc, A. M. (2003). The friendship features of preschool children: Links with prosocial behavior and aggression. *Social Development, 12,* 249–268. doi: 10.1111/1467-9507.00232

Shure, M., & Spivack, G. (1982). Interpersonal problem solving in young children: A cognitive approach to prevention. *American Journal of Community Psychology, 10,* 341–356.

Smetana, J. G. & Ball, C. (2018). Young children's moral judgments, justifications, and emotion attributions in peer relationship contexts. *Child Development, 89,* 2245-2263. doi: 10.1111/cdev.12846

Smetana, J. G. & Braeges, J. L. (1990). The development of toddlers' moral and conventional judgments. *Merrill-Palmer Quarterly, 36,* 329–346.

Smetana, J. G., Campione-Barr, N., & Yell, N. (2003). Children's moral and affective judgments regarding provocation and retaliation. *Merrill-Palmer Quarterly, 49,* 209–237. doi: 10.1353/mpq.2003.0010

Smetana, J. G., Jambon, M., & Ball, C. (2014). The social domain approach to children's moral and social judgments. In M. Killen & J. G. Smetana (Eds), *Handbook of moral development* (2nd ed., pp. 23-45). doi: 10.4324/9780203581957.ch2

Smetana, J. G., Jambon, M., & Ball, C. (2018). Normative changes and individual differences in early moral judgments: A constructivist developmental perspective. *Human Development, 64,* 264-280. doi: 10.1159/000492803

Smetana, J. G., Jambon, M., Conry-Murray, C., & Sturge-Apple, M. L. (2012). Reciprocal associations between young children's developing moral judgements and theory of mind. *Developmental Psychology, 48,* 1144-1155. doi: 10.1037/a0025891

Smetana, J. G., Toth, S. L., Cicchetti, D., Bruce, J., Kane, P., & Daddis, C. (1999). Maltreated and nonmaltreated preschoolers' conceptions of hypothetical and actual moral transgressions. *Developmental Psychology, 35,* 269–281. doi: 10.1037/0012-1649.35.1.269

Spivak, A. L. (2016). Dynamics of young children's socially adaptive resolutions of peer conflict. *Social Development, 25*, 212–232. doi: 10.1111/sode.12135

Stifter, C. A., Cipriano, E., Conway, A., & Kelleher, R. (2009). Temperament and the development of conscience: The moderating role of effortful control. *Social Development, 18*, 353-374. doi: 10.1111/j.1467-9507.2008.00491.x

Stifter, C. A., Putnam, S., & Jahromi, L. (2008). Exuberant and inhibited toddlers: Stability of temperament and risk for problem behavior. *Development and Psychopathology, 20*, 401–421. doi: 10.1017/S0954579408000199

Thompson, S. F., Zalewski, M., Kiff, C. J., Moran, L., Cortes, R., & Lengua, L. J. (2020). An empirical test of the model of socialization of emotion: Maternal and child contributors to preschoolers' emotion knowledge and adjustment. *Developmental Psychology, 56*, 418–430. doi: 10.1037/dev0000860

ToonDoo. (2012). www.toondon.com. Jambav, Pleasonton, CA.

Tremblay, R. E. (2010). Developmental origins of disruptive behaviour problems: the "original sin" hypothesis, epigenetics and their consequences for prevention. *Journal of Child Psychology and Psychiatry, 51*, 341–367. doi: 10.1111/j.1469-7610.2010.02211.x

Tremblay, R. E., Nagin, D. S., Seguin, J. R., Zoccolillo, M., Zelazo, P. D., Boivin, M., … & Japel, C. (2004). Physical aggression during early childhood: Trajectories and predictors. *Clinical Neurophysiology, 123*, 106–116. doi: 10.1016/j.clinph.2011.06.006.A

Utendale, W. T., Hubert, M., Saint-Pierre, A. B., & Hastings, P. D. (2011). Neurocognitive development and externalizing problems: The role of inhibitory control deficits from 4 to 6 years. *Aggressive Behavior, 37*, 476–488. doi: 10.1002/ab.20403

Vaillancourt, T., Miller, J. L., Fagbemi, J., Côté, S., & Tremblay, R. E. (2007). Trajectories and predictors of indirect aggression: results from a nationally representative longitudinal study of Canadian children aged 2-10. *Aggressive Behavior, 33*, 314–326. doi: 10.1002/ab.20202

Vera-Estay, E., Seni, A. G., Champagne, C., & Beauchamp, M. H. (2016). All for one: Contributions of age, socioeconomic factors, executive functioning, and social cognition to moral reasoning in childhood. *Frontiers in Psychology, 7*, 1–13. doi: 10.3389/fpsyg.2016.00227

Vitaro, F., Barker, E. D., Boivin, M., Brendgen, M., & Tremblay, R. E. (2006). Do early difficult temperament and harsh parenting differentially predict reactive and proactive aggression?. *Journal of Abnormal Child Psychology, 34*, 685–695. doi: 10.1007/s10802-006-9055-6

Waltzer, T., Baxley, C., & Dahl, A. (2019). Adults' responses to young children's transgressions: A new method for understanding everyday social interactions. *Early Child Development and Care,* 1-15. doi: 10.1080/03004430.2019.1709182

Walker, O. L., & Henderson, H. A. (2012). Temperament and social problem solving competence in preschool: Influences on academic skills in early elementary school. *Social Development, 21*, 761–779. doi: 10.1111/j.1467-9507.2011.00653.x

Webster-Stratton, C., Reid, M., & Stoolmiller, M. (2008). Preventing conduct problems and improving school readiness: Evaluation of the Incredible Years teacher and child training programs in high risk schools. *Child Psychology and Psychiatry, 49,* 471–488. doi: 10.1111/j.1469-7610.2007.01861.x

Wechsler, D. (2003).*Wechsler Intelligence Scale for Children—4th Edition* (WISC-IV).San Antonio, TX: Harcourt Assessment.

White, B. A., Jarrett, M. A., & Ollendick, T. H. (2013). Self-regulation deficits explain the link between reactive aggression and internalizing and externalizing behavior problems in children. *Journal of Psychopathology and Behavioral Assessment, 35*, 1–9. doi: 10.1007/s10862-012-9310-9

Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): A method of assessing executive function in children. *Nature Protocols, 1,* 297–301. doi: 10.1038/nprot.2006.46

Zelazo, P. D. (2015). Executive function: Reflection, iterative reprocessing, complexity, and the developing brain. *Developmental Review, 38,* 55–68. doi: 10.1016/j.dr.2015.07.001

Zelazo, P. D., Anderson, J. E., Richler, J., Wallner-Allen, K., Beaumont, J. L., & Weintraub, S. (2013). II. NIH Toolbox Cognition Battery (Cb): Measuring executive function and attention. *Monographs of the Society for Research in Child Development, 78*(4), 16–33. doi: 10.1111/mono.12032

Zelazo, P. D., & Carlson, S. M. (2012). Hot and cool executive function in childhood and adolescence: Development and plasticity. *Child Development Perspectives, 6,* 354–360. doi:10.1111/j.1750-8606.2012.00246.x

Zelazo, P. D., Helwig, C., & Lau, A. (1996). Intention, act, and outcome in behavioral prediction and moral judgment. *Child Development, 65*, 2478–2492. doi:10.2307/1131635

Zelli, A., Dodge, K. A., Lochman, J. E., Laird, R. D., & Conduct Problems Prevention Research Group. (1999). The distinction between beliefs legitimizing aggression and deviant processing of social cues: Testing measurement validity and the hypothesis that biased processing mediates the effects of beliefs on aggression. *Journal of Personality and Social Psychology, 77,* 150–166. doi: 10.1037/0022-3514.77.1.150

Zeman, J., Shipman, K., & Suveg, C. (2002). Anger and sadness regulation: Predictions to internalizing and externalizing symptoms in children. *Journal of Clinical Child and Adolescent Psychology, 31*, 393–398. doi: 10.1207/153744202760082658

Zhou, Q., Chen, S. H., & Main, A. (2012). Commonalities and differences in the research on children's effortful control and executive function: A call for an integrated model of self-regulation. *Child Development Perspectives, 6*, 112–121. doi: 10.1111/j.1750-8606.2011.00176.x

APPENDIX A: LIST OF SITUATIONS ON THE VSG

Situation	Situation Description	Situation Type
Physical Aggression	Transgressor kicks the participant while playing in the sandbox	Conflict
Product Destruction	Transgressor knocks over participant's block tower	Conflict
Product Inequity	Transgressor takes away the participant's soccer ball	Conflict
Social Exclusion	Transgressor says that he will not play with the participant, but will play with another child instead	Conflict
Teasing (ability)	Transgressor says the participant's drawing of a dog looks like an ugly monster and laughs	Conflict
Teasing (personal choice)	Transgressor laughs at the participant's doll and says "you're a baby"	Conflict
Greeting	A classmate waves hi to the participant	Benign
Offer Resources	A classmate offers the participant a pencil while drawing	Benign
Play Request	A classmate asks the participant to play with him	Benign

.

APPENDIX B: SOCIAL INTERVIEW VIGNETTES AND INTERVIEW QUESTIONS

Vignettes:

Ripping Art Work (product destruction): Madison/Mason(transgressors), Sarah/Sam (victims)

One day, Madison and Sarah were coloring next to each other. Madison leaned over, took Sarah's picture, and ripped it in half.

After Madison ripped Sarah's picture, Sarah …

- … hit Madison (physical aggression).
- … called Madison names and said "you better not do that again or else!" (verbal aggression).
- … told Madison "you can't play with me anymore!" (relational aggression).

Taking Toy (product inequity): Jenna/Jason, Tina/Tommy

One day, Jenna and Tina were playing with toy cars. While Tina wasn't looking, Jenna took Tina's car, leaving Tina with no cars to play with.

When Tina saw that Jenna took her car, she …

- … hit Jenna (physical aggression).
- … called Jenna names and said "you better not do that again or else!" (verbal aggression).
- … told Jenna "you can't play with me anymore!" (relational aggression).

Knocking over house (product destruction): Caitlin/Caleb, Megan/Matthew

One day, Caitlin and Megan were building houses for their stuffed animals to live in. Megan came up to Caitlin and knocked over the house Caitlin had just built.

After Megan knocked over her house, Caitlin …

- … hit Megan (physical aggression).
- … called Megan names and said "you better not do that again or else!" (verbal aggression).
- … told Megan "you can't play with me anymore!" (relational aggression).

Interview Questions:

Example vignette response: Sarah hits Madison after Madison ripped her picture

1. **Act acceptability:** How good or bad is it for Sarah to hit Madison after Madison ripped Sarah's picture?

1- Very good 2- A little good 3- Just ok 4- A little bad 5- Very bad

2. **Deserved punishment:** What do you think should happen to Sarah after she hits Madison? Should Sarah get …

1- A big prize, like a toy 2- A small prize, like a sticker 3- Nothing
4- In a little trouble 5- In a lot of trouble

3. **Authority independence**: What if the teacher saw Sarah hit Madison and says, "I don't care that you hit Madison." If it was okay with the teacher, how good or bad was it for Sarah to hit Madison after Madison ripped Sarah's picture?

1- Very good 2- A little good 3- Just ok 4- A little bad 5- Very bad

4. **Rule independence:** What if there was no classroom rule about hitting and it was ok to hit in the classroom? If it was not breaking any rules, how good or bad was it for Sarah to hit Madison after Madison ripped Sarah's picture?

1- Very good 2- A little good 3- Just ok 4- A little bad 5- Very bad

.

APPENDIX C: CBQ-VSF ITEMS-BY-SCALE

Surgency
1 Seems always in a big hurry to get from one place to another.
4 Likes going down high slides or other adventurous activities.
7 Often rushes into new situations.
10 Seems to be at ease with almost any person.
13R Prefers quiet activities to active games.
16 Likes to go high and fast when pushed on a swing.
19R Takes a long time in approaching new situations.
22R Is sometimes shy even around people s/he has known a long time.
25 Is full of energy, even in the evening.
28 Likes rough and rowdy games.
31R Is slow and unhurried in deciding what to do next.
34R Sometimes turns away shyly from new acquaintances.

Negative Affect
2 Gets quite frustrated when prevented from doing something s/he wants to do.
5 Is quite upset by a little cut or bruise.
8 Tends to become sad if the family's plans don't work out.
11 Is afraid of burglars or the "boogie man."
14 When angry about something, s/he tends to stay upset for ten minutes or longer.
17 Seems to feel depressed when unable to accomplish some task.
20R Hardly ever complains when ill with a cold.
23 Is very difficult to soothe when s/he has become upset.
26R Is not afraid of the dark.
29R Is not very upset at minor cuts or bruises.
32 Gets angry when s/he can't find something s/he wants to play with.
35 Becomes upset when loved relatives or friends are getting ready to leave following a visit.

Effortful Control
3 When drawing or coloring in a book, shows strong concentration.
6 Prepares for trips and outings by planning things s/he will need.
9 Likes being sung to.
12 Notices it when parents are wearing new clothing.
15 When building or putting something together, becomes very involved in what s/he is doing, and works for long periods.
18 Is good at following instructions.
21 Likes the sound of words, as in nursery rhymes.
24 Is quickly aware of some new item in the living room

27 Sometimes becomes absorbed in a picture book and looks at it for a long time.
30 Approaches places s/he has been told are dangerous slowly and cautiously.
33 Enjoys gentle rhythmic activities, such as rocking or swaying.
36 Comments when a parent has changed his/her appearance.

.

APPENDIX D: SCBE-30 AND PTOPS

SOCIAL COMPETENCE AND BEHAVIOR EVALUATION SHORT FORM

Instructions: Please indicate how frequently this child engages in each listed behavior.

Use the following scale to answer:
Circle 1 if this behavior almost never occurs
Circle 2 if this behavior rarely occurs
Circle 3 if this behavior sometimes occurs
Circle 4 if this behavior often occurs
Circle 5 if this behavior frequently occurs
Circle 6 if this behavior almost always occurs

(Items are grouped by subscale for clarity, but will be randomized when presented to the teacher)

Anxiety/Withdrawn						
Neutral facial expression	1	2	3	4	5	6
Fearfulness	1	2	3	4	5	6
Sad/depressed	1	2	3	4	5	6
Inhibited in groups	1	2	3	4	5	6
Inactive/prefers to watch activities	1	2	3	4	5	6
Isolated	1	2	3	4	5	6
Tired	1	2	3	4	5	6
Does not interact in groups	1	2	3	4	5	6
Unnoticed in groups	1	2	3	4	5	6
Worries	1	2	3	4	5	6
Anger/Aggression						
Frustrated	1	2	3	4	5	6
Angry when interrupted	1	2	3	4	5	6
Irritable	1	2	3	4	5	6
Screams/yells	1	2	3	4	5	6
Hits/bites/kicks	1	2	3	4	5	6

Forces others to do unwanted things	1	2	3	4	5	6
Hits/destroys things	1	2	3	4	5	6
Defiant when reprimanded	1	2	3	4	5	6
Oppositional	1	2	3	4	5	6
<u>Social Competence</u> Comforts/assists others	1	2	3	4	5	6
Helpful with routine tasks	1	2	3	4	5	6
Works well in groups	1	2	3	4	5	6
Uses negotiation	1	2	3	4	5	6
Considers other's points	1	2	3	4	5	6
Cooperates with other children	1	2	3	4	5	6
Takes care of toys	1	2	3	4	5	6
Attends to younger children	1	2	3	4	5	6
Accepts reasonable compromise	1	2	3	4	5	6
Happy with own accomplishments	1	2	3	4	5	6

PRESCHOOL TAXONOMY OF PROBLEM SITUATIONS

Instructions: For each situation, please indicate how likely this child is to respond in an inappropriate manner (e.g., by hitting peers, aggressing verbally, crying, disrupting the group, withdrawing, appealing to the teacher for help, or behaving in some other immature, unacceptable, and unsuccessful way). In other words, how much of a problem is this situation for the child?

Use the following scale to answer:
Circle 1 if this situation is never a problem for this child.
Circle 2 if this situation is rarely a problem for this child.
Circle 3 if this situation is sometimes a problem for this child.
Circle 4 if this situation is usually a problem for this child.
Circle 5 if this situation is almost always a problem for this child.

For example: When this child is teased by peers

If you feel that when this child is teased by peers, he or she almost always responds inappropriately or ineffectively (such as by crying), you would agree that this is a problem situation for this child and would circle 5. If you feel that when this situation occurs, the child almost always responds in an effective and appropriate manner (such as by ignoring the teasing), you would agree that this is not a problem situation for this child and would circle 1. We are less interested in how frequently this situation occurs and more interested in this child's response when it does occur.

1. When a peer takes this child's turn during an activity with established procedures for turn-taking.	1	2	3	4	5
2. When peers call this child a bad name.	1	2	3	4	5
3. When a peer has something belonging to this child, and this child wants it back.	1	2	3	4	5
4. When this child has something belonging to a peer and the peer wants it back before this child is through with it.	1	2	3	4	5
5. When this child is playing with a peer, and the peer accidentally breaks this child's toy.	1	2	3	4	5
6. When this child is teased by peers.	1	2	3	4	5
7. When this child is accidentally provoked by a peer (such as a peer who accidentally bumps into this child in line).	1	2	3	4	5
8. When a peer has a toy or object that this child wants.	1	2	3	4	5
9. When a peer expresses anger at this child.	1	2	3	4	5

10. When a peer takes a turn in place of this child.	1	2	3	4	5
11. When this child has been teased or threatened, s/he gets angry easily and strikes back.	1	2	3	4	5
12. This child always claims that other children are to blame in a fight and feels that they started the trouble.	1	2	3	4	5
13. When a peer accidentally hurts this child (such as by bumping into him), s/he overreacts with anger and fighting.	1	2	3	4	5
14. When a peer refuses to play with this child, s/he gets angry and threatens the peer.	1	2	3	4	5
15. When a peer takes an object from this child, s/he gets angry and will use force to retrieve the object.	1	2	3	4	5
16. When this child makes a request of a peer and the peer refuses, this child gets angry and either threatens the peer or strikes out at the peer.	1	2	3	4	5
17. When a peer ignores this child, s/he gets angry and either threatens the peer or strikes out at the peer.	1	2	3	4	5
18. When a peer refuses to play with this child, s/he gets angry and either threatens the peer or strikes out at the peer.	1	2	3	4	5