Foundational Skills among Black and Latino Children in Poverty: Predicting School-Related Outcomes and Exploring Family Stress Processes as Correlates of Development

A dissertation
presented to the faculty of the
Curry School of Education
University of Virginia

in partial fulfillment
of the requirements for the degree
Doctor of Philosophy in Education

by
Chelsea A. K. Duran

August 2018

Funding: The research reported here was supported by awards from the Institute of Education Sciences, U.S. Department of Education, through Grant #R305A110703 to the University of Virginia. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Department of Educational Leadership, Foundations, and Policy
Educational Psychology – Applied Developmental Science
Curry School of Education
University of Virginia
Charlottesville, VA

APPROVAL OF THE DISSERTATION

The dissertation ("Foundational Skills among Black and Latino Children in Poverty: Predicting School-Related Outcomes and Exploring Family Stress Processes as Correlates of Development") has been approved by the Graduate Faculty of the Curry School of Education in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

____
Name of Chair (Dr. David W. Grissmer)

____
Committee Member (Dr. Jessica Whittaker)

____
Committee Member (Dr. Daphna Bassok)

____
Committee Member (Dr. Erik Ruzek)

Date: _________________________

Table of Contents

CONCEPTUAL LINK: Introduction to the Three Manuscripts – Foundational Skills among Black and Latino Children in Poverty: Predicting School-Related Outcomes and Exploring Family Stress Processes as Correlates of Development

In an era of rising economic inequality and persistent racial inequity, the long-term opportunities for Black and Latino children living in poverty in the United States (U.S.) are increasingly constrained as a function of their family background (Duncan & Murnane, 2011; Ladson-Billings, 2006; Putnam, 2016; Reardon, Robinson, & Weathers, 2015). Public education exists, in part, to equalize opportunities by providing *all* children, regardless of their family background, quality learning experiences supporting later economic and general well-being (*Brown v. Board of Education*, 1954; U.S. Department of Health and Human Services, 2001). Therefore, diminishing societal inequity as a function of family backgrounds is within the mission of public education system.

Historically, White children from middle- to upper-income families have constituted the majority among the student population in the U.S. education system and among participants of developmental research. As a result, current educational systems and processes (e.g., instructional techniques) may be built upon assumptions regarding children's "normative" development which are inconsistent with the development of children outside of this population (Delpit, 2006; Garcia-Coll et al., 1996). Improving equity, then, may require better understanding how cognitive skills and behavioral tendencies, especially those systematically varying with family background but not explicitly taught or tested in school, relate to academic outcomes in non-White and/or low-income populations. Given that distal contextual factors, including structural inequalities in society, tend to affect children through interactions within their most proximal contexts, and their families in particular (Blair & Raver, 2012; Bronfenbrenner, 1986; Conger et al., 2002; McLoyd, 1998; Mistry et al., 2002), it is also important to understand how family processes relate to children's development in these populations.

Motivated by this need, the three papers in this dissertation focus on foundational, developmental outcomes among mostly Black and Latino children from low-income families during the first three years of elementary school, including their relations with school-related outcomes and family stress processes. The first identifies unique, statistical contributions of skills in the domains of self-regulation and visuo-spatial ability to children's mathematics learning over the course of one school year. Guided by a causal model for how economic hardship affects family stress and children's development, the second paper in this dissertation focuses on relations between multiple family stress processes and the development of two self-regulatory outcomes (i.e., executive functions [EFs] and delay of gratification). The third investigates the psychometric properties of a delay of gratification measure, including its relations with school-related outcomes. Insights from these papers may inform, for example, prevention programs targeting low-income families (Blair & Raver, 2016; Neville et al., 2013), as well as programs targeting non-academic skills which are currently underway in many schools serving high proportions of non-White children from low-income families (Leving, 2014; McNeel, 2013; Moreton, 2014).

Poverty, Inequality, Ubiquitous Instability, and Race-based Tension as the Backdrop for Study Children's Development

The studies in this dissertation are based on data derived from two broader studies. The first, from which data for Paper 1 are derived, tested an intervention within the context of an after-school social and emotional learning (SEL) program. The second, from which data for Paper 2 and Paper 3 are derived, was an experimental evaluation of the SEL program itself.

Children and their families were recruited to participate in each study from several elementary schools within the same, urban school district in the Southeastern U.S. This section provides an account of the community characteristics, many the result of entrenched poverty and systemic racism. While these broader community characteristics outside of family life are not an explicit focus of the present studies, they constitute the backdrop for participating children's development.

Deep poverty and racial inequality are prevalent features of this community and correspond to inequity in children's educational and behavioral outcomes. In the county in which study children lived at the start of the second broader study (from which data for papers 2 and 3 are derived), more than one quarter of children were living below the federal poverty threshold (kidscount.org; specific citation suppressed). Black and Hispanic/Latino children in the county were 10 and five to eight times as likely, respectively, as White children to live in areas of concentrated poverty (kidscount.org; specific citation suppressed). The characteristics of the specific neighborhoods served by study schools are mirrored in the demographic make-up of the study samples, with close to 90% of children being Black, 5% Hispanic/Latino, and about 95% receiving free- or reduced-price lunch (an indicator of low-income status). The racial homogeneity of, and lack of representation of White children in, the sample is striking, given that the district overall serves roughly the same percentage (45%) of White and Black children. This fact speaks to the extent of residential and school segregation in the county in line with norms across Southern states (Frankenberg, Hawley, Ee, & Orfield, 2017; Suitts, 2016).

The high proportion of poverty rates among families in this community are also reflected in familial socioeconomic indicators in the sample traditionally thought to reflect families' capacity to invest in children's learning and development. Twenty-nine percent of mothers had less than a high school diploma, for instance, and about a quarter were pregnant as teenagers. Further, more than half of study children (54%) were living with only one parent (usually, their mothers). While these characteristics are considered in the literature to be indicative of low-resource and high-stress home environments, the resilience of caregivers (most of whom were mothers) is evident in their low reports of financial strain (see descriptive statistics in Paper 2). Caregivers' resilience was further expressed in (unpublished) qualitative interviews, where many exhibited an ability to reframe life stressors as opportunities (e.g., losing a job as an opportunity which led to a better job). And so, despite traditional socioeconomic indicators suggesting most study families had low capacity to financially invest in their children, these indicators do not capture the emotional capacity of these families to support their children in the face of financial and social hardship.

In children's other most proximal environments – their schools and classrooms – quality of children's classroom interactions was lower than national norms (West & Adams, 2016; West, Adams, & Kim, 2017). This is consistent with the fact that, generally speaking, children in urban districts from low-income families tend to be taught by lower skilled teachers (Lankford, Loeb, & Wyckoff, 2002). It is not surprising, then that study participants underperformed on academic achievement measures relative to children from other studies of higher SES. For example, the children in the present studies scored close to 2 standard deviations lower, on average, on an applied problem solving (mathematics) measure than the middle-SES participants in Cameron and colleagues' (2012) study when measured at the start of kindergarten. This gap narrowed substantially, by about a third, by the end of the kindergarten year, but remained above a full standard deviation difference.

In the broader community encompassing the school district, community cohesion and race relations and community cohesion were strained. At the start of the evaluation study, crime rates in the county were among the highest in the country (citation suppressed) and, during the course of the same study, two high-profile incidents of race-based violence occurred in the area. Each received national attention (citations suppressed), and one incident – which involved the killing of an unarmed Black man by a White police officer – played a role in fueling the national, emerging Black Lives Matter movement (citation suppressed). These specific events resounded with the history of racial tensions and racism in the county and surrounding region (citation suppressed). Elevated crime and violence rates are common in areas marked by concentrated poverty, perhaps best explained by the role poverty plays in breaking down social cohesion and trust within the community (Sharkey, Besbris, and Friedson, 2016). Thus, violence in the area, as result of both entrenched poverty and long-standing strain in race relations, may be a manifestation of low collective trust in these communities. Social cohesion and trust in the area is perhaps especially low between racial groups and between historically oppressed minority groups and dominant social institutions. This is a potentially important point for understanding the classroom context for children, as a majority of study children were taught by White teachers.

In short, the children who participated in the present studies faced immense hardship, both in terms of low economic resources, high instability, and racial tension. Being of predominantly racial/ethnic minority status and from impoverished communities, they represent a population whose development is not well understood and has historically not been well regarded by the education system in the U.S. Thus, in the context of intensifying concern regarding family background as a constraint on children's academic opportunities, it is important to understand the development and relevance of psychological constructs not traditionally explicitly addressed in school in this particular population.

Differences in Foundational Skills May Contribute to Achievement Gaps

Non-White children from low-income families have often been labeled as “at-risk” for school difficulties, at least in part due to differences in certain developmental skills between these children and White children from middle- to upper-income families (Evans & Rosenbaum, 2008; Ursache, Blair, & Raver, 2012; Verdine, Golinkoff et al., 2014). Thus, there has been considerable interest in the early childhood education community in identifying specific skills which support children’s early learning and adaptation to school, and which differ systematically by SES. Two relevant domains of development considered in this dissertation set are self-regulation and perceptuo-motor skills.

Self-regulation. Broadly defined as the intentional direction and modification of behavior in service of goals or in line with context-specific rules or norms, self-regulation supports positive adaptation in a wide range of life domains (Blair & Raver, 2012; Moffit et al., 2011), including school and academics (Brock et al., 2009; McClelland & Cameron, 2011; Ursache, Blair, & Raver, 2012). Classroom environments differ from the typical home or preschool environment in several ways, including more intentional adult-child interactions focused on progress toward academic goals, longer periods of attention, and the need to get along with larger, more diverse groups of peers (Rimm-Kaufman & Pianta, 2000). Success in this environment relies heavily on children’s ability to regulate their thoughts, emotions, and behavior, which may be one reason outcomes in this domain predict academic performance and learning (McClelland & Cameron, 2011). Self-regulatory outcomes tend to differ between children of different socioeconomic status (SES), as described further below. These observations have motivated the advocacy of promoting self-regulatory strategies which are adaptive in school environments as a promising direction in efforts to improve the academic prospects of children from low-SES families (Ursache et al., 2012).

Executive Functions (EFs). EFs – or the cognitive processes underpinning self-regulation, including attention, inhibitory control, working memory, and cognitive flexibility – are closely related to children's transition to school and to later academic and life outcomes from preschool through adulthood (Blair & Razza, 2007; McClelland, Acock, Piccinin, Rhea, & Stallings, 2013; McClelland et al., 2007; Willoughby, Blair, Wirth, & Greenberg, 2012). Theoretically, these relations reflect EFs' foundational role in fulfilling the behavioral and cognitive demands of school context (Blair, 2002; McClelland & Cameron, 2011). For instance, raising one's hand before speaking requires inhibiting the urge to speak out immediately, and flexibly switching problem-solving strategies on different mathematics problems requires cognitive flexibility. Specific EFs, as well as general EF ability, in early childhood are empirically linked and may be especially important for mathematics achievement (Bull, Epsy, & Wiebe, 2008; Bull & Scerif, 2001; Clark, Sheffield, Wiebe, & Epsy, 2013; Cragg & Gilmore, 2014; Geary, 2011; LeFevre et al., 2014). Empirical research has also demonstrated positive correlations between SES and EF development (Hackman, Gallop, Evans, & Farah, 2015; Noble, Norman, & Farah, 2005). This link may, at least partially, mediate the positive correlation between SES and academic outcomes (Fitzpatrick, McKinnon, Blair, & Willoughby, 2014) including mathematics outcomes in particular (Dilworth-Bart, 2012).

Delay of gratification. Delay of gratification refers to the tendency to choose later rewards over less valuable, immediate rewards. When measured in early childhood, delay of gratification is a robust predictor of welfare in many domains of adult life, including educational outcomes (also, health, finances, incarceration; Casey et al., 2011; Duckworth, 2011; Evans, Fuller-Rowell, & Doan, 2012; Mischel, 2014; Moffit et al., 2011), even after controlling for potential confounds such as intelligence and SES (Duckworth & Seligman, 2005; Moffit et al., 2011). Delay of gratification at age 4.5 years partially mediated the relation between income-to-needs ratio and grade 5 academic achievement in a large, nationally representative sample (Evans & Rosenbaum, 2008), and delay of gratification in late elementary school fully mediated class differences in middle school course grades in a mostly White, rural sample (Evans & English, 2002). However, the strength of relations between SES-related factors and delay of gratification, as well as between delay of gratification and school-related outcomes, has been found most often using measures tapping into not only children's preferences – or choices – to delay gratification, but also their ability to sustain their self-imposed choice. These relations are less clear-cut in research utilizing measures focused only on the choice to delay gratification.

Whether delay of gratification operates similarly among Black and Latino children from low-income families, as compared to the way it operates among White children from relatively affluent families, is not clear. Not only are children of color traditionally underrepresented in the research on which current theory is based (see Moreton, 2014), but certain cognitive processes are differentially correlated with delay of gratification performance between racial groups within low-income samples (Price-Williams & Ramirez, 1974; Strickland, 1972). Clarifying the nature of delay of gratification development specifically within the population represented by the present studies, then, may help to refine delay of gratification as a construct. It may also provide important insight for emerging efforts to promote social and emotional development in schools, which often target similar populations and often include delay of gratification-related outcomes in their list of desired skills and behaviors (Moreton, 2014).

Perceptuo-motor skills. Other skills linked to academic achievement include the coordination of perceptual and motor abilities. Such skills may be directly implicated in common classroom activities (e.g., tracing and copying letters, measuring objects) and therefore support academic functioning in general. Visuo-spatial abilities in particular may also undergird specific quantitative and mathematical learning, such as the development of a mental number line (Dehaene, 2011; Dehaene & Cohen, 2007; Geary, 2001).

Visuo-motor integration (VMI). Visuomotor integration (VMI) – closely related to the more colloquially-known "hand-eye coordination" – requires several component skills, such as visual attention and motor control (Beery & Beery, 2004). VMI in early childhood significantly predicts later academic achievement, and improvement in achievement, especially in mathematics (Dinehart & Manfra, 2013). This relation remains even after controlling for EF and other cognitive abilities in middle-SES (Cameron et al., 2012) and nationally representative samples (Grissmer et al., 2010). Visuospatial skills among children from low-income households tend to be under-developed relative to those of their more affluent counterparts (Potter, Mashburn, & Grissmer, 2013). Thus, it may be important to understand the extent to which such abilities contribute to academic achievement among low-SES children.

Paper 1: Unique and Compensatory Associations of Executive Functioning and Visuomotor Integration with Mathematics Performance in Early Elementary School

The first paper in this dissertation set has contributed to the literature focused on relations between developmental skills and academic achievement in racial/ethnic minority and low-income samples. Specifically, I examined unique associations between EFs and VMI with mathematics performance, and improvement in multiple measures of mathematics performance, among kindergarteners (n = 89, M_{age} = 5.5 years) and first graders (n = 73, M_{age} = 6.6 years).

Consistent with theoretical and unique roles each skill plays in supporting mathematics performance and learning, both EFs and VMI were associated with both concurrent performance and improvement in performance on most individual mathematics assessments included in this study. Results did not support a compensatory pattern, or any other type of interaction effect, for these cognitive skills in relating to most mathematics outcomes, except perhaps for two measures which draw heavily on visuo-spatial ability and informal mathematics skills.

Paper 2: Family Stress Processes and Self-Regulatory Development of Black and Hispanic/Latino Children from Low-income Families

In addition to identifying foundational skills for learning among racial/ethnic minority children from low-income families, it is important to understand the nature of foundational skills' development in order to properly interpret children's skills and behaviors. In early childhood, development may be most profoundly affected by interactions and processes within children's family, which relay effects from more distal contexts, such as neighborhoods or parental workplaces (Bronfenbrenner, 1986). Poverty and economic hardship are associated with high family stress (Conger et al., 2002; Mistry et al., 2006) and family instability (Sturge-Apple et al., 2017), which may influence children's self-regulatory development. Thus, paper 2 in this dissertation set focuses on the development of two self-regulatory outcomes – EFs and delay of gratification – as they relate to these processes.

Children's self-regulatory abilities develop adaptively based on experiences, especially those in early childhood (Blair & Raver, 2012), given rapid development during that timeframe (Best, Miller, & Jones, 2009; Lengua et al., 2015). Home environments and caregivers play an influential role in 'canalizing' self-regulatory tendencies. For instance, families with ample and consistent financial resources tend to reinforce "reflective modes" of self-regulation, characterized by self-restraint, future-oriented thinking/planning, and preference toward delaying gratification. Other home environments, especially those with scarce or inconsistent resources, may reinforce “reflexive modes,” characterized by quick thinking and pre-emptive behaviors and, perhaps, preferences toward immediate gratification. While a reactive mode may be beneficial in high stress or instable environments, due in part to lower certainty of future events (Blair & Raver, 2012; Ellis, Bianchi, Griskevicius, & Frankenhuis, 2017), related tendencies can lead to disadvantages in academic experiences, as schools largely favor a reflective mode of self-regulation (Finegood & Blair, 2017). Scarcity and unpredictability in life of may be two contextual factors leading to the development of a reflexive self-regulatory mode, which is more common among children from low-SES families (Belsky, Schlomer, & Ellis, 2012; Blair & Raver, 2012; Blair et al, 2015). Thus, self-regulation may represent a domain in which a mismatch exists between the actual development of low-SES children and the assumptions implicitly held within schools regarding this development.

Caregiving and stress are likely mechanistic pathways mediating relations between contextual hardship and self-regulatory outcomes (Blair et al., 2011; Evans & Kim, 2013; Hackman, Farah, & Meaney, 2010; Lengua, Honorado, & Bush, 2007; Lengua et al., 2014). The widely validated Family Stress Model (FSM) describes how economic hardship can affect children’s development, ultimately through child-caregiver interactions (Conger et al., 2002). Guided by the FSM, the study presented in Paper 2 examined direct and indirect relations from multiple family stress processes, measured at two time points, to children’s improvement in EF and delay of gratification from the beginning of kindergarten to the beginning of first grade (ages 5 – 7; n = 343). Pathways from family stress processes to children’s self-regulation differed by outcome. A mediating role for child-caregiver interactions between other family stress processes and children's outcomes (as in the FSM) was replicated in the EF model, but not in the delay of gratification model. Instead, only financial strain – theoretically a more distal stress process than child-caregiver conflict – exhibited significant relations with delay of gratification development. Further, only changes in family stress processes, as opposed to baseline levels in each, were directly related to each outcome. Taken together, these results underscore child-caregiver interactions as a worthy target for intervention. But, they also reveal diversity in pathways from family stress to children's development based on specific outcome and may suggest that addressing immediate consequences of economic hardship on caregivers may be necessary for alleviating some of poverty's effects on children.

Paper 3: Validation of a Choice Delay of Gratification Measure among Children of Color from Low-Income Families

Delay of gratification tasks have long been used as measures of self-control, and in some forms show strong predictive relations with outcomes much later in life (e.g., Casey et al., 2011; Moffit et al., 2011). Validation studies of "choice" delay of gratification measures in particular have yielded little evidence for their tie to other self-regulatory outcomes (e.g., Hongwanishkul, Happaney, Lee, & Zelazo, 2005), but have been carried out almost exclusively in White, middle- to upper-income samples. In light of emerging theories regarding delay of gratification as differentially adaptive based on context (Ellis et al., 2017), the present study examines the psychometric properties of a choice delay measure in a sample of mostly Black (some Latino/a) kindergarteners (n = 354; M_{age} = 5.4 years) attending an urban school district serving predominantly low-income families.

In a categorical confirmatory factor analysis with a single factor, the measure exhibited adequate internal reliability and invariance across relevant subgroups. However, choice delay performance in this sample was negatively correlated with other concurrently measured, self-regulatory constructs. Further, choice delay performance positively predicted difficulties transitioning to school year and problem classroom behaviors, negatively predicted positive classroom behaviors, and did not predict academic achievement. Sensitivity analyses testing the predictive validity of alternative scoring schemes for choice delay performance revealed similar patterns across schemes and outcomes. These results either call into question the validity of choice delay tasks as self-control or necessitate a more nuanced theoretical interpretation of delay of gratification behaviors in children of color and/or in the context of poverty.

Common Strengths and Limitations

The papers in this study share key strengths and limitations. Each paper includes multiple time points of data, thereby strengthening each respective body of empirical, which often utilize cross-sectional research designs. Similarly, each paper simultaneously considers relations among multiple independent variables with dependent variables – each of which may be confounded in others if studied separately. While the analytic approach of each paper is still observational, each builds a stronger case for the current, causal theories posed as explanations for relations than that provided by other existing bodies of research. Often, prior studies are based on cross-sectional data and/or do not include key confounding variables in analysis, and so leave open a higher probability of alternative explanations for their results.

Being of predominantly racial/ethnic minority status and from impoverished communities, the children who participated in the present studies are representative of a population historically under-represented in research and under-served by school systems in the U.S. As such, it is important to study the nature of foundational, developmental skills in this and similar samples. Such inquiry may lead to insights regarding how schooling experiences may be better designed to meet the needs and strengths of an ever-diversifying student population.

Conclusion and Implications

Responding to a growing body of evidence suggesting nonacademic skills are important for academic success (SES; Heckman & Kautz, 2012; Blair & Diamond, 2008), educational organizations have engaged in efforts to explicitly foster self-control, among other non-academic skills, in low-SES children and youth (McNeel, 2013). Such programs seek to foster what is often referred to as “character development,” or a collection of skills – including delay of gratification and, more generally, self-control – statistically related to academic success in the literature. Historically speaking, these programs represent a progressive step forward, in that they take on a strengths-oriented – as opposed to a traditional, deficit-oriented – approach to supporting at-risk children and youth. At the same time, “character development” initiatives have been criticized for promoting the development of an arbitrarily-chosen “bag of virtues.” The shape and content of these efforts may be heavily influenced by psychological theory and research (Moreton, 2014), but are being applied in populations that have been under-represented in this research base. As the functional adaptivity of such non-academic skills is not well understood in populations of color and/or the context of poverty (Ellis et al., 2017; Wadsworth, 2015), framing them in terms of “virtues” or “character” may be unnecessarily stigmatizing.

More research is needed to clarify whether the same skills related to academic achievement among White and/or middle- to upper-SES children are also important for children outside of this historically “normative” population. Consideration of out-of-school influences on development of foundational skills may also inform how and why deviation from developmental “norms” might be adaptive in children’s homes and neighborhoods. An understanding of development as a process of adaptation may lead to a mental shift from viewing children who face adversity as weakened or needing "fixing," to one which views children as optimally developed for thriving in their most proximal contexts (Ellis et al., 2017). When such adaptive development is at-odds with what is expected by society's socially dominant institutions, such as schools (Finegood & Blair, 2017), disparaging children's home environments and experiences may need to be intentionally avoided.

References

Banks, W. C., McQuater, G. V., Ross, J. A., & Ward, W. E. (1983). Delayed gratification in blacks: A critical review. *Journal of Black Psychology*, *9*(2), 43-56.

Beery, K. E., & Beery, N. A. (2004). The *Beery-Buktenica developmental test of visual-motor integration* manual. Bloomington, MN: Pearson Assessments.

Belsky, J., Schlomer, G. L., & Ellis, B. J. (2012). Beyond cumulative risk: distinguishing harshness and unpredictability as determinants of parenting and early life history strategy. *Developmental psychology*, *48*(3), 662.

Best, J. R., Miller, P. H., & Jones, L. L. (2009). Executive functions after age 5: Changes and correlates. *Developmental review*, *29*(3), 180-200.

Blair, C. (2002). School readiness: Integrating cognition and emotion in a neurobiological conceptualization of children's functioning at school entry. *American psychologist*, *57*(2), 111.

Blair, C., & Diamond, A. (2008). Biological processes in prevention and intervention: The promotion of self-regulation as a means of preventing school failure. *Development and psychopathology*, *20*(3), 899-911.

Blair, C., Granger, D. A., Willoughby, M., Mills-Koonce, R., Cox, M., Greenberg, M. T., ... & Fortunato, C. K. (2011). Salivary cortisol mediates effects of poverty and parenting on executive functions in early childhood. *Child development*, *82*(6), 1970-1984.

Blair, C., & Raver, C. C. (2012). Child development in the context of adversity: experiential canalization of brain and behavior. American Psychologist, 67(4), 309.

Blair, C., & Raver, C. C. (2016). Poverty, stress, and brain development: New directions for prevention and intervention. *Academic pediatrics*, *16*(3), S30-S36.

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child development*, *78*(2), 647-663.

Brock, L. L., Rimm-Kaufman, S. E., Nathanson, L., & Grimm, K. J. (2009). The contributions of 'hot'and 'cool'executive function to children's academic achievement, learning-related behaviors, and engagement in kindergarten. *Early Childhood Research Quarterly*, *24*(3), 337-349.

Bronfenbrenner, U. (1986). Ecology of the family as a context for human development: Research perspectives. *Developmental psychology*, *22*(6), 723.

Brown v. Board of Education, 347 U.S. 483 (1954).

Bull, R., Espy, K. A., & Wiebe, S. A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. Developmental neuropsychology, 33(3), 205-228.

Bull, R., & Scerif, G. (2001). Executive functioning as a predictor of children's mathematics ability: Inhibition, switching, and working memory. *Developmental neuropsychology*, *19*(3), 273-293.

Cameron, C. E., Brock, L. L., Murrah, W. M., Bell, L. H., Worzalla, S. L., Grissmer, D., & Morrison, F. J. (2012). Fine motor skills and executive function both contribute to kindergarten achievement. Child development, 83(4), 1229-1244.

Casey, B. J., Somerville, L. H., Gotlib, I. H., Ayduk, O., Franklin, N. T., Askren, M. K., ... & Glover, G. (2011). Behavioral and neural correlates of delay of gratification 40 years later. *Proceedings of the National Academy of Sciences*, *108*(36), 14998-15003.

Clark, C. A., Sheffield, T. D., Wiebe, S. A., & Espy, K. A. (2013). Longitudinal associations between executive control and developing mathematical competence in preschool boys and girls. *Child Development*, *84*(2), 662-677.

Conger, R. D., Wallace, L. E., Sun, Y., Simons, R. L., McLoyd, V. C., & Brody, G. H. (2002). Economic pressure in African American families: a replication and extension of the family stress model. *Developmental psychology*, *38*(2), 179.

Cragg, L., & Gilmore, C. (2014). Skills underlying mathematics: The role of executive function in the development of mathematics proficiency. *Trends in Neuroscience and Education*, *3*(2), 63-68.

Dehaene, S., & Cohen, L. (2007). Cultural recycling of cortical maps. Neuron, 56(2), 384-398.

Dinehart, L., & Manfra, L. (2013). Associations Between Low-Income Children's Fine Motor Skills in Preschool and Academic Performance in Second Grade. Early Education & Development, 24(2), 138-161.

Dehaene, S. (2011). The number sense: How the mind creates mathematics. Oxford University Press.

Dehaene, S., & Cohen, L. (2007). Cultural recycling of cortical maps. Neuron, 56(2), 384-398.

Delpit, L. (2006). *Other people's children: Cultural conflict in the classroom*. The New Press.

Duckworth, A. L. (2011). The significance of self-control. Proceedings of the National Academy of Sciences, 108(7), 2639-2640.

Duckworth, A. L., & Seligman, M. E. (2005). Self-discipline outdoes IQ in predicting academic performance of adolescents. Psychological science, 16(12), 939-944.

Dilworth-Bart, J. E. (2012). Does executive function mediate SES and home quality associations with academic readiness?. *Early Childhood Research Quarterly*, *27*(3), 416-425.

Duncan, G. J., & Murnane, R. J. (Eds.). (2011). *Whither opportunity?: Rising inequality, schools, and children's life chances*. Russell Sage Foundation.

Ellis, B. J., Bianchi, J., Griskevicius, V., & Frankenhuis, W. E. (2017). Beyond risk and protective factors: An adaptation-based approach to resilience. *Perspectives on Psychological Science*, *12*(4), 561-587.

Evans, G. W. (2004). The environment of childhood poverty. American psychologist, 59(2), 77.

Evans, G. W., & English, K. (2002). The environment of poverty: Multiple stressor exposure, psychophysiological stress, and socioemotional adjustment. *Child development*, *73*(4), 1238-1248.

Evans, G. W., Fuller-Rowell, T. E., & Doan, S. N. (2012). Childhood cumulative risk and obesity: The mediating role of self-regulatory ability. *Pediatrics*, *129*(1), e68-e73.

Evans, G. W., & Kim, P. (2013). Childhood Poverty, Chronic Stress, Self-Regulation, and Coping. *Child Development Perspectives, 7*, 43-48. doi:10.1111/cdep.12013

Evans, G. W., & Rosenbaum, J. (2008). Self-regulation and the income-achievement gap. *Early Childhood Research Quarterly*, *23*, 504-514. doi:10.1016/j.ecresq.2008.07.002

Finegood, E. D. & Blair, C. (2017). Poverty, Parent Stress, and Emerging Executive Functions in Young Children. In Deater-Deckar, K. & Panneton, R. (Eds.) *Parental Stress and Early Child Development: Adaptive and Maladaptive Outcomes.* Springer International Publishing AG, Switzerland

Fitzpatrick, C., McKinnon, R. D., Blair, C. B., & Willoughby, M. T. (2014). Do preschool executive function skills explain the school readiness gap between advantaged and disadvantaged children? *Learning and Instruction*, *30*, 25-31.

Frankenberg, E., Hawley, G. S., Ee, J., & Orfield, G. (2017). Southern schools: more than a half-century after the civil rights movement. *The Civil Rights Project*. Penn State Center for Education and Civil Rights, University Park, PA. Accessed April 8, 2018: https://www.civilrightsproject.ucla.edu/research/k-12-education/integration-and-diversity/southern-schools-brown-83-report/Brown63_South_052317-RELEASE-VERSION.pdf

Garcia Coll, C., Crnic, K., Lamberty, G., Wasik, B. H., Jenkins, R., Garcia, H. V., & McAdoo, H. P. (1996). An integrative model for the study of developmental competencies in minority children. Child development, 67(5), 1891-1914.

Gates, S. M., Ringel, J. S., Santibanez, L., Guarino, C., Ghosh-Dastidar, B., & Brown, A. (2006). Mobility and turnover among school principals. *Economics of Education Review*, *25*(3), 289-302.

Geary, D. C. (2001). A Darwinian Perspective on Mathematics and Instruction. In Loveless, T. (ed.; 2001). The Great Curriculum Debate: How Should We Teach Reading and Math?" The Brookings Institution, Washington D.C.

Geary, D. C. (2011). Cognitive predictors of achievement growth in mathematics: a 5-year longitudinal study. *Developmental psychology*, *47*(6), 1539.

Grissmer, D., Grimm, K. J., Aiyer, S. M., Murrah, W. M., & Steele, J. S. (2010). Fine motor skills and early comprehension of the world: two new school readiness indicators. Developmental psychology, 46(5), 1008.

Hackman, D. A., Farah, M. J., & Meaney, M. J. (2010). Socioeconomic status and the brain: mechanistic insights from human and animal research. *Nature Reviews Neuroscience*, *11*(9), 651.

Hackman, D. A., Gallop, R., Evans, G. W., & Farah, M. J. (2015). Socioeconomic status and executive function: Developmental trajectories and mediation. *Developmental Science*, *18*(5), 686-702.

Heckman, J. J., & Kautz, T. (2012). Hard evidence on soft skills. *Labour economics*, *19*(4), 451-464.

Hongwanishkul, D., Happaney, K. R., Lee, W. S., & Zelazo, P. D. (2005). Assessment of hot and cool executive function in young children: Age-related changes and individual differences. *Developmental neuropsychology*, *28*(2), 617-644.

Ladson-Billings, G. (2006). From the achievement gap to the education debt: Understanding achievement in US schools. *Educational researcher*, *35*(7), 3-12.

Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: A descriptive analysis. *Education Evaluation and Policy Analysis*.

LeFevre, J. A., Fast, L., Skwarchuk, S. L., Smith-Chant, B. L., Bisanz, J., Kamawar, D., & Penner-Wilger, M. (2010). Pathways to mathematics: Longitudinal predictors of performance. *Child development*, *81*(6), 1753-1767.

Lengua, L. J., Honorado, E., & Bush, N. R. (2007). Contextual risk and parenting as predictors of effortful control and social competence in preschool children. *Journal of applied developmental psychology*, *28*(1), 40-55.

Lengua, L. J., Kiff, C., Moran, L., Zalewski, M., Thompson, S., Cortes, R., & Ruberry, E. (2014). Parenting mediates the effects of income and cumulative risk on the development of effortful control. *Social Development*, *23*(3), 631-649.

Lengua, L. J., Moran, L., Zalewski, M., Ruberry, E., Kiff, C., & Thompson, S. (2015). Relations of growth in effortful control to family income, cumulative risk, and adjustment in preschool-age children. *Journal of abnormal child psychology*, *43*(4), 705-720.

Leving, D. (2014). "Dave Levin on Not Eating the Marshmallow." Knowledge is Power Program (KIPP) Blog. http://blog.kipp.org/teamandfamily/dave-levin-on-not-eating-the-marshmallow/

Marcynyszyn, L. A., Evans, G. W., & Eckenrode, J. (2008). Family instability during early and middle adolescence. *Journal of Applied Developmental Psychology*, *29*(5), 380-392.

McClelland, M. M., Acock, A. C., Piccinin, A., Rhea, S. A., & Stallings, M. C. (2013). Relations between preschool attention span-persistence and age 25 educational outcomes. *Early childhood research quarterly*, *28*(2), 314-324.

McClelland, M. M., & Cameron, C. E. (2011). Self-regulation and academic achievement in elementary school children. *New Directions for Child and Adolescent Development*, *2011*(133), 29-44.

McClelland, M. M., Cameron, C. E., Connor, C. M., Farris, C. L., Jewkes, A. M., & Morrison, F. J. (2007). Links between behavioral regulation and preschoolers' literacy, vocabulary, and math skills. *Developmental psychology*, *43*, 947.

McLoyd, V. C. (1998). Socioeconomic disadvantage and child development. *American psychologist*, *53*(2), 185.

McNeel, B. (2013) "KIPP Inspires Inner-City Students: 'Don't Eat the Marshmallow Yet'." *Rivard Report.* http://therivardreport.com/kipp-inspires-inner-city-children-dont-eat-the-marshmallow-yet/

Mistry, R. S., Vandewater, E. A., Huston, A. C., & McLoyd, V. C. (2002). Economic well-being and children's social adjustment: The role of family process in an ethnically diverse low-income sample. *Child development*, *73*(3), 935-951.

Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H. L., Houts, R., Poulton, R., Roberts, B.W., Ross, S. and Sears, M.R. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. Proceedings of the National Academy of Sciences, 108(7), 2693-2698.

Moore, K. A., Vandivere, S., & Ehrle, J. (2000). Turbulence and Child Well-Being. New Federalism: National Survey of America's Families, Series B, No. B-16. Assessing the New Federalism: An Urban Institute Program To Assess Changing Social Policies.

Moreton, B. (2014). S'More Inequality The Neoliberal Marshmallow and the Corporate Reform of Education. Social Text, 32(3 120), 29-48.

Neville, H. J., Stevens, C., Pakulak, E., Bell, T. A., Fanning, J., Klein, S., & Isbell, E. (2013). Family-based training program improves brain function, cognition, and behavior in lower socioeconomic status preschoolers. *Proceedings of the National Academy of Sciences*, *110*(29), 12138-12143.

Noble, K. G., Norman, M. F., & Farah, M. J. (2005). Neurocognitive correlates of socioeconomic status in kindergarten children. *Developmental science*, *8*(1), 74-87.

Price-Williams, D. R., & Ramirez III, M. (1974). Ethnic differences in delay of gratification. The Journal of Social Psychology, 93(1), 23-30.

Ross, L. T., & Hill, E. M. (2002). Childhood unpredictability, schemas for unpredictability, and risk taking. *Social Behavior and Personality: an international journal*, *30*(5), 453-473.

Strickland, B. R. (1972). Delay of gratification as a function of race of the experimenter. *Journal of personality and Social Psychology*, *22*(1), 108.

Sturge-Apple, M. L., Davies, P. T., Cicchetti, D., Hentges, R. F., & Coe, J. L. (2017). Family instability and children's effortful control in the context of poverty: Sometimes a bird in the hand is worth two in the bush. *Development and psychopathology*, *29*(3), 685-696.

Suitts, S. (2016). Students Facing Poverty: The New Majority. *Educational Leadership*, *74*(3), 36-40.

Potter, D., Mashburn, A., & Grissmer, D. (2013). The family, neuroscience, and academic skills: An interdisciplinary account of social class gaps in children's test scores. *Social Science Research*, *42*(2), 446-464.

Putnam, R. D. (2016). *Our kids: The American dream in crisis*. Simon and Schuster.

Reardon, S.F., Robinson, J.P., & Weathers, E.S. (2015). Patterns and Trends in Racial/Ethnic and Socioeconomic Academic Achievement Gaps. In H. A. Ladd & E. B. Fiske (Eds.), *Handbook of Research in Education Finance and Policy (Second ed.)*. Lawrence Erlbaum.

Rimm-Kaufman, S. E., & Pianta, R. C. (2000). An ecological perspective on the transition to kindergarten: A theoretical framework to guide empirical research. *Journal of applied developmental psychology*, *21*(5), 491-511.

Sharkey, P., Besbris, M., & Friedson, M. (2016). Poverty and Crime. In *The Oxford Handbook of the Social Science of Poverty*. Accessed April 8, 2018: http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199914050.001.0001/oxf ordhb-9780199914050-e-28

Suitss, S. (2016). Race and Ethnicity in a New Era of Public Funding of Private Schools: Private School Enrollment in the South and the Nation. *Atlanta: Southern Education Foundation, Inc*.

Tobin, R. M., & Graziano, W. G. (2010). Delay of Gratification: A Review of Fifty Years of Regulation Research. In Hoyle, R. H. (Ed.) (2013). Handbook of Personality and Self-regulation. John Wiley & Sons. pp. 47-63.

U.S. Census Bureau, 2008-2012, 2009-2013, 2010-2014, 2011-2015,and 2012-2016 American Community Survey 5-Year Estimates, Tables B01001, B01001B-B01001H, and B17001.

United States Department of Health and Human Services (2001). "The Economic Rationale for investing in children: A Focus on Child Care. Rationale for Govt Role in Education. Dec 1, 2001" United States Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation, Washington, D. C. Last Accessed April 7, 2018: https://aspe.hhs.gov/report/economic-rationale-investing-children-focus-child-care/rationale-government-role-education

Ursache, A., Blair, C., & Raver, C. C. (2012). The promotion of self-regulation as a means of enhancing school readiness and early achievement in children at risk for school failure. *Child Development Perspectives*, *6*(2), 122-128.

Verdine, B. N., Golinkoff, R. M., Hirsh-Pasek, K., Newcombe, N. S., Filipowicz, A. T., & Chang, A. (2014). Deconstructing building blocks: Preschoolers' spatial assembly performance relates to early mathematical skills. *Child development*, *85*(3), 1062-1076.

Wadsworth, M. E. (2015). Development of maladaptive coping: A functional adaptation to chronic, uncontrollable stress. *Child development perspectives*, *9*(2), 96-100.

West, H., Adams, C.L., & Kim, H. (2017, April). Classroom quality as a predictor of classroom behavior in a sample of socio-demographically at-risk students. Poster session at the 2017 meeting of the Society for Research in Child Development (SRCD), April 6th-8th, Austin, Texas.

West, H. & Adams, C.L. (2016, April). Classroom organization as a predictor of academic achievement in kindergarten. Paper session at the 2016 meeting of the Southeastern Psychological Association (SEPA), March 30th-April 2nd, New Orleans, Louisiana.

Willoughby, M. T., Blair, C. B., Wirth, R. J., & Greenberg, M. (2012). The measurement of executive function at age 5: Psychometric properties and relationship to academic achievement. *Psychological assessment*, *24*(1), 226.

MANUSCRIPT ONE: Unique and Compensatory Associations of Executive Functioning and Visuomotor Integration with Mathematics Performance in Early Elementary School

Chelsea A.K. Duran[1], Anthony Byers[2], Claire E. Cameron[3], and David Grissmer[1]

1. Center for Advanced Study of Teaching and Learning, Curry School of Education, University of Virginia, PO Box 800784, Charlottesville, VA 22908-0784, USA.

2. Graham and Parks School, 44 Linnaean St. Cambridge, MA 02138, USA.

3. Graduate School of Education, University at Buffalo, The State University of New York, 572 Baldy Hall, Buffalo, NY 14260-1000 USA.

Citation:

Duran, C. A., Byers, A., Cameron, C. E., & Grissmer, D. (2018). Unique and compensatory associations of executive functioning and visuomotor integration with mathematics performance in early elementary school. *Early Childhood Research Quarterly*, *42*, 21-30.

Acknowledgements:
The authors sincerely thank the schools, teachers, and families who participated in this research and without whom this study would not have been possible.
Funding: The research reported here was supported by awards from the Institute of Education Sciences, U.S. Department of Education, through Grant #R305A110703 to the University of Virginia. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Abstract

Research has illuminated contributions—usually modeled separately—of both executive functioning (EF) and visuomotor integration (VMI) to mathematical development in early elementary school. This study examined simultaneous associations of EF and VMI, measured at the beginning of the school year, with concurrent and later mathematics performance on several mathematics assessments in kindergartners (n = 89, M_{age} = 5.5 years) and first graders (n = 73, M_{age} = 6.6 years) of low socioeconomic status. Both skills were related to concurrent performance on all assessments, as well as improvement through the end of the school year for all but a geometry subtest, which was predicted only by VMI. No significant influence of an interaction between the skills was present, except for concurrently on the geometry subtest and longitudinally on an assessment with a relatively strong emphasis on informal skills. Findings are discussed in the context of supporting mathematics development in early childhood.

Keywords: Executive Functioning; Visuomotor Integration; Mathematics; Cognitive Development; Low Socioeconomic Status

Unique and Compensatory Associations of Executive Functioning and Visuomotor Integration with Mathematics Performance in Early Elementary School

Introduction and Literature Review

Children's early mathematics ability is an important factor supporting academic success (e.g., Duncan et al., 2007), but other less explicitly academic skills, when measured in early childhood, are also strongly associated with later academic outcomes (Heckman, Stixrud, & Urzua, 2006). Executive functioning (EF) and visuomotor integration (VMI) upon school entry are two such skills, especially in the way they support mathematics learning (Cragg & Gilmore, 2014; Verdine, Irwin, Golinkoff, & Hirsh-Pasek, 2014b). Each of these skills shows consistent associations with later mathematics outcomes (e.g., Blair & Razza, 2007; Newcombe & Frick, 2010) and theory has independently been developed for each skill regarding why each might support academic achievement and mathematics achievement in particular. However, despite these skills appearing to co-develop in early childhood (see Cameron, Cottone, Murrah, & Grissmer, 2016), studies linking them to mathematics performance have often considered them in isolation. Thus, whether each skill makes a unique contribution to mathematics performance when the other is measured is less well understood, as is whether strengths in one skill might compensate for relative weaknesses in the other. Even less understood is whether such contributions are consistent across different types of mathematics assessments.

Understanding the contributions of cognitive skills to academic outcomes may be especially important in populations with low socioeconomic status (SES; i.e., racial minority status and/or of low-income), who are traditionally considered at-risk for school difficulties. This heightened risk can, in part, be explained by fewer opportunities for children of low socioeconomic status to fully develop the cognitive skills which promote academic adjustment and performance (Bradley & Caldwell, 2013; Verdine et al., 2014a).

This study examines both EF and VMI as predictors of concurrent mathematics performance and improvement in several measures of mathematics performance over the course of one school year in a sample of mostly African-American kindergarteners and first graders from low-income families. Based on research with preschool-age children (Cameron et al., 2015), compensatory effects – by which strengths in one skill counteract effects of short-comings in the other – between EF and VMI were also tested.

EF and Academic Performance

EF is a cognitive construct typically described as the higher cognitive processes underlying conscious control of thought and action (Jacques & Marcovitch, 2010). These processes fall into correlated but distinct components: working memory, cognitive flexibility, and inhibitory control (Diamond, 2013; Miyake et al., 2000). In this study, EF was measured using a composite from multiple tasks.

Development in EF is characterized by several periods of rapid growth, particularly during early childhood (Diamond, 2006; Romine & Reynolds, 2005). Early overall EF, as well as specific components of EF, are linked to a wide range of academic and life outcomes from preschool through adulthood (Diamond, 2013; McClelland, Acock, Piccinin, Rhea, & Stallings, 2013). For instance, McClelland and colleagues (2007) found that incoming EF in preschool was correlated with initial academic skills but also predicted preschoolers' math and literacy improvements by the end of preschool. EF is also associated with academic achievement in math and literacy after children transition to kindergarten (Cameron Ponitz, McClelland, et al., 2009), even after controlling for IQ (Blair & Razza, 2007). The association between preschool EF and academic skills is consistent at least through college (Duncan et al., 2007; McClelland et al., 2013).

Beyond facilitating appropriate classroom engagement, EF supports specific academic tasks, especially in mathematics (Geary, 2011; Willoughby, Blair, Wirth, & Greenberg, 2012). For example, children with low inhibitory control – one component of EF – may be less likely to evaluate and switch mathematical problem-solving strategies when they prove ineffective (Bull & Scerif, 2001). EF seems to be particularly important for word problems, which require students to build and manipulate models of the problems in their heads (Fuchs et al., 2010ab).

Empirical evidence confirms that EF undergirds performance on mathematical tasks, with many studies demonstrating their concurrent and longitudinal links. For instance, both overall EF and specific components of EF are known to longitudinally predict mathematics performance in early childhood (Clark, Sheffield, Wiebe, & Epsy, 2013), early elementary school (Geary, 2011), and middle elementary school (LeFevre et al., 2014). Numerous studies demonstrate that EF, and especially working memory, accounts for significant variance in mathematics performance above and beyond other potential confounds, including prior mathematics performance, language skills, and even IQ (see Cragg & Gilmore, 2014 for a review). Recent studies have even suggested that EF may, at least partially, mediate the relationship between SES and mathematics outcomes (Dilworth-Bart, 2012) and researchers have suggested that promoting EF as part of encouraging children to develop adaptive self-regulatory strategies could be a viable means to improve the academic outcomes of children from low-SES families (Ursache, Blair, & Raver, 2012).

VMI and Academic Performance

VMI is considered an aspect of visuospatial processing and, like EF, is a complex process requiring integration of multiple skills, namely, visual and motor functioning (Beery &

Buktenica, 1997). The most prominently used measures of VMI are design-copying tasks, in which individuals are typically asked to copy a series of increasingly complex figures. While copying a design may seem like a fairly simple task, it requires several component skills, chief among them the ability to parse a whole figure into its parts and to reintegrate those parts into a whole (Akshoomoff & Stiles, 1995).

The development of VMI, which is preceded by development of more rudimentary visual and motor skills prior to their integration develops rapidly during early childhood (ages 4 through 7 years), but continues through at least age 12 (Decker, Englund, Carboni, & Brooks, 2011). Past research drew attention to visuospatial skills in general, and VMI in particular, by documenting links between early “visuospatial” or “fine motor” skills and later academic achievement (e.g., Cameron et al., 2012; Carlson, Rowe, & Curby, 2013; Grissmer, Grimm, Aiyer, Murrah, & Steele, 2010; Luo, Jose, Huntsinger, & Pigott, 2007; Stevenson & Newman, 1986). These studies’ measures often included design-copying tasks, which as mentioned above are elsewhere and in this study are considered measures of VMI (e.g., Beery & Buktenica, 1997, Korkman, Kirk, & Kemp, 1998). As discussed next, several studies document links of VMI with mathematics performance, specifically.

Developmental theory and research has long highlighted the notion that basic mathematical competencies (e.g., concepts of number) are directly dependent upon visuospatial development. Indeed, mathematical tasks in the classroom (e.g., estimating the number of objects in a jar, measuring the length of an object, copying diagrams from the board) often require visuospatial skills and VMI. Further, research in cognitive neuroscience points to a neural basis for the link between visuospatial skills and mathematical abilities. Dehaene and Cohen (2007) suggest that the neural machinery supporting humans’ sophisticated visual spatial processing abilities may be “recycled” to form powerful visual representations of mathematical concepts (e.g., as in the development of a mental number line; Dehaene, 2011).

Empirical evidence has also documented links between visuospatial skills and mathematics development in children. In two studies of early elementary school students, one with a low-SES sample, and the other with a middle-SES sample, Gunderson, Ramirez, Beilock, and Levine (2012) found that children’s visuospatial skills predicted later number line knowledge and calculation skills, even after controlling for prior number line knowledge and math ability. While these findings are what one might expect given work linking visuospatial skills with the mental number line in adults (e.g. Doricchi, Guariglia, Gasparini, & Tomaiuolo, 2005), it is also possible that EF was a confound in these relationships, as it was not measured. Although visuospatial skills and EF are distinct cognitive processes, any task that requires attentional resources and problem solving will also utilize EF (Korkman et al., 1998), and visuospatial skills tasks are no exception. Indeed, at least one study has found that the association between VMI and mathematics does not hold after controlling for certain measures of attention, which are considered an aspect of EF (Sortor & Kulp, 2003). Therefore, it is important to measure EF when examining visuospatial (and VMI) skills’ unique contribution to early math skills, and vice versa. Further, because children from low-income households tend to have nascent visuospatial skills early in elementary school (Potter, Mashburn, & Grissmer, 2013), it may be particularly important to understand to what extent this ability contributes to mathematics performance in a sample of children from low-SES families.

Unique Contributions of EF and VMI to Mathematics Performance

EF and visuospatial skills in general, and VMI in particular, may exhibit co-dependency in development and may interact in nuanced ways to support children’s adaptation in school

(Cameron et al., 2016; Assel, Landry, Swank, Smith, & Steelman, 2003). Ample evidence exists to support the notion that EF is an important promotive factor for school success and mathematics performance in particular, above and beyond other potential confounds, such as intelligence (see Cragg & Gilmore, 2014 for a review). Less research has focused on whether apparent contributions of visuospatial ability in general, and VMI in particular, to mathematics performance remain after accounting for other cognitive skills such as EF. Thus, while there is reason to suspect that EF should predict mathematics performance even when accounting for visuospatial skills, such as VMI, it is less clear that the converse is true (i.e., that VMI should predict mathematics performance after accounting for EF).

Two studies utilizing teacher reports for EF found unique associations between visuospatial skills and mathematics performance. Assel and colleagues (2003) measured visuospatial skills, in part, with a design-copying task and EF using teacher report at ages three, four, and six; and mathematics ability at age eight in low- to lower-middle-SES children. They found that both predictors were highly correlated from one time point to another (i.e., visuospatial skills to visuospatial skills and EF to EF) and that visuospatial skills and teacher-rated EF at age six were each uniquely associated with mathematics ability at age eight. The study did not include a measure of mathematical ability prior to age 8, however. Similarly, Grissmer and colleagues (2010) analyzed six nationally representative data sets, including ECLS-K, and found that age 5 measures of both VMI and teacher-reported EF predicted math ability in fifth grade, even after controlling for mathematics performance at kindergarten entry. Thus, both studies suggest that visuospatial skills are important above and beyond their relationship with EF, though measurement of EF in these studies was through teacher-report, which is conceptually different from directly-measured EF, as in this study.

Findings from research using direct measures of EF during the transition to formal schooling have sometimes, but not always, been consistent with research using teacher-reports, described immediately above. For example, in a small (n = 44), economically diverse sample of preschoolers, direct measures of both EF and visuospatial skills – including VMI – were uniquely related to concurrent mathematics (Verdine et al., 2014b). In slight contrast, Cameron and colleagues (2012) found in a middle- to upper-middle-SES sample that kindergarten-entry EF, but not VMI, was robustly associated with initial mathematics as well as to gains in mathematics. Sample demographics including age may explain these discrepancies: Cameron et al.’s (2012) sample was mostly middle- to upper-middle-class children who have, on average, stronger visuospatial skills at school entry than their lower SES peers (Potter et al., 2013). Thus, many of the children in the Cameron et al. (2012) study may already have had a level of visuospatial skills that readied them for kindergarten mathematics, which underscores the need for research examining the role of visuospatial skills in low-SES children. Finally, Becker, Miao, Duncan, and McClelland (2014) found both EF and VMI to predict later performance on applied problem solving mathematics assessment in an economically diverse sample of preschool and kindergarten-aged children. Verdine and colleagues (2014b) and Becker and colleagues (2014) did not include a pretest for mathematics performance, however, and therefore whether both skills would uniquely predict mathematics performance above and beyond prior performance is unclear. Furthermore, both Cameron and colleagues (2012) and Becker and colleagues (2014) focused on only one type of mathematics assessment (i.e., applied problem solving), leaving open the possibility that relations between EF and visuospatial skills and concurrent and later mathematics might be different for other types of mathematics (e.g., geometry). The present study examined both concurrent and longitudinal relations between EF, VMI, and multiple measures of mathematics performance in a sample of low-SES children, in order to explore whether EF and VMI predict improvement across a variety of types of mathematics performance in this population during early elementary school.

Potential for a compensatory effect of EF and visuospatial skills. There is a nuanced relationship between EF and visuospatial skills (Cameron et al., 2016) due in part to the fact that both draw from the same limited cognitive resource (Floyer-Lea & Matthews, 2004). This relationship appears to have implications for academic functioning: for example, in one low-SES sample of 4-year-olds, interaction effects between aspects of EF and VMI were found such that good VMI compensated for low inhibitory control in predicting gains in academic outcomes, including emergent literacy and behavior (Cameron et al., 2015). In other words, gains in these outcomes for children with low inhibitory control were similar to those with strong inhibitory control if they exhibited strong VMI ability. However, others have not replicated this result in elementary school for outcomes such as classroom behaviors (Kim et al., 2016). Thus, this study explores the potential that compensatory interaction effects between EF and VMI may also exist in the current, low-SES sample for a variety of mathematics performance measures, for which such interactions have not been investigated.

Present Study

This study had two broad aims. The first aim was to determine, in a low-SES sample, the unique contribution of EF and VMI to children’s concurrent early mathematics performance at the beginning of the school year (Time 1) and to improvement in their mathematics performance through the end of the school year (Time 2, controlling for Time 1). This study was largely expected to replicate previous findings; specifically, both EF and VMI were expected to significantly predict mathematics performance concurrently (at Time 1, as in Verdine et al., 2014 and Cameron et al., 2012) and longitudinally (at Time 2, controlling for Time 1, as in Cameron et al., 2012 and Becker et al., 2014).

The second aim was to explore a potential compensatory association between EF and VS on concurrent mathematics performance and improvement in mathematics performance. In other words, we sought to investigate whether EF and visuospatial skills additively predicted mathematics, or whether strengths in only one skill sufficed for strong mathematics performance. Such compensatory effects have been found for literacy outcomes in a low-SES preschool sample (Cameron et al., 2015), though questions remain about whether this pattern will emerge for mathematics measured in elementary school.

Method

The present, observational and exploratory study, is based on data from an experimental study of an intervention targeting fine motor skills and delivered in the context of an afterschool, social and emotional learning program serving three urban elementary schools in a large city in the southeastern United States. Children were recruited from their schools before being randomized into either a treatment condition (offered opportunity to receive intervention) or a control condition (not offered opportunity to receive intervention). The intervention was administered over the course of one school year. Only the intervention delivered to treatment children in the first cohort was found to have effects; these were on attention and visuospatial skills, but not mathematics (Grissmer et al., 2013). The present study is observational and exploratory in nature, in that it has a correlational design – without strong causal warrant – and utilizes data which were ultimately collected for a different purpose. With respect to the second study aim in particular, we do not have firm hypotheses regarding whether compensatory effects may be more or less prominent for specific types of mathematics. Sensitivity analyses, described further below, were conducted to test whether the participation in the intervention may have interfered with our study results.

Participants

Participants were 162 children in kindergarten (*n* = 89) and first grade (*n* = 73) who were recruited in two successive cohorts of 97 and 65 children, respectively. Children were assessed at the start (Time 1) and end (Time 2) of the school year, before and after delivery of the aforementioned intervention. Average age at Time 1 was 5.5 years (SD_{months} = 4.0) for kindergarteners and 6.6 years (SD_{months} = 4.5) for first graders; average time between Time 1 and Time 2 assessments was 6.5 months. All children in the afterschool program were eligible to participate in the broader study, with the exception of those with severe disabilities that would prevent them from completing the assessment battery. Of the 162 children included in this study, 50% were boys, 92% were African American, and 95% qualified for free or reduced-price lunch, which roughly reflects the demographic characteristics of the student population served by the schools from which students were recruited.

Measures

EF and VMI were tested using the Developmental Neuropsychological Assessment (NEPSY; Korkman et al., 1998), a neuropsychological battery normed for ages 3-12 years. The NEPSY comprises subtests that measure five broad neuropsychological domains, including Attention/Executive Functions and Visuospatial Processing. Herein we provide internal and stability reliability information for these measures, when possible. For some, we include only stability reliability coefficients (i.e., test-retest coefficients between Time 1 and Time 2 assessments) because for these tests internal reliability is not appropriate due to dependency between item-level scores and/or multi-dimensionality in the scores (e.g., as in the use of both time and accuracy in the scoring criteria; Korkman et al., 1998).

Executive Functioning. The NEPSY Attention/EF domain score comprised three subtests: Tower, Auditory Attention and Response Set, and Visual Attention. On Tower, children were shown pictures of three colored balls on three pegs and asked to copy the pattern in a set number of moves using a physical set of three balls and pegs. Correctly-executed sequences within a time limit were scored as correct. There were 20 items, and thus possible raw scores were 0 to 20. The test was discontinued if and when a child incorrectly responded to four consecutive items. Tasks similar to Tower (e.g., Tower of London, Tower of Hanoi) have elsewhere been described as broad or complex EF tasks, in that they require executive planning and problem solving, as opposed to only tapping a single component of EF, such as working memory or inhibitory control (Culbertson & Zilmer, 1998; Miyake et al., 2000). Internal reliability (Cronbach's alpha) for Tower was .84 at Time 1 and .88 for Time 2, and test-retest reliability was $r = .54$.

On Auditory Attention and Response Set, children heard recorded lists of color and non-color words in two trials. On trial 1, children were instructed to place a colored piece of foam in a box when they heard that color. The rules were changed for trial 2, such that responses depended on which color word was heard. Depending on the word, instructions dictated a word color and foam color match, mismatch, or no response. For example, the word "yellow" indicated that children should put a red piece of foam in the box, and the word "red" indicated that they should put a yellow piece of foam in the box. Correct responses were those that followed the rule in under 3 seconds (higher scores were given for quicker responses). This task required that children utilize their attention and working memory to act in accordance with the instructions and rules. Because rules changed between trial 1 and trial 2, this task also required inhibitory control and cognitive flexibility in order to inhibit performing according to previously used rules. The test-retest reliability of the Auditory Attention and Response Set subtest in this sample was $r = .49$.

On Visual Attention, children chose a target picture (trial 1) or pictures (trial 2) from an array of similar pictures presented on paper. Responses were scored as correct if children circled the target pictures in the array within a set time limit of 3 minutes. Performance on this task reflects a child's ability to focus selectively on and maintain attention to visual targets in the process of a visual search. The test-retest reliability of the Visual Attention subtest scaled scores in this sample was r = .30.

EF core domain scores, which are meant to represent a child's performance relative to a normed population and, were calculated according to NEPSY guidelines (Korkman et al., 1998), were used as the final measure of EF.). Core domain scores take into account all subtests within each domain, consider child age, and are scaled so that, in the normative sample, the mean is 100 and standard deviation is 15 points. The core domain test-retest reliability scores for the Attention/Executive domain was r = .60.

Visuomotor Integration. VMI was measured using the Design Copying subtest of the NEPSY Visuospatial Processing domain. On this test, children used paper and pencil to copy two-dimensional geometrical designs of increasing complexity. Designs were scored according to established criteria, with up to four points awarded on each of 18 items and, hence, a maximum score of 72. The test was discontinued if and when a child incorrectly responded to four consecutive items. Internal reliability (Cronbach's alpha) for Design Copying in this sample was 0.79 at Time 1 and .72 at Time 2; test-retest reliability was r = .72.

Mathematics Performance. Mathematics performance was assessed using direct assessments from three different validated test batteries: the Woodcock-Johnson III (WJ III) Tests of Achievement (Woodcock, McGrew, Mather, & Schrank, 2001), KeyMath3-3 (Connolly, 2007), and the Test of Early Mathematics Ability-3 (TEMA; Ginsburg & Baroody, 2003). Composite mathematics scores were calculated at each time point by averaging the *z*-scores for each test.

Woodcock-Johnson Applied Problems. The WJ III Applied Problems subtest assesses the ability to analyze and solve math problems in context. Because the problems are presented within multiple modalities (pictures, word problems, etc.), children must ignore extraneous information, recognize which procedure to use, and then perform the appropriate calculations. For example, children were shown a picture of five cookies and were asked, “If Jessica ate three of these cookies, how many cookies are left?” The ability to solve later items requires specific knowledge, such as coin denominations and how to read a clock. In this sample, Cronbach’s alpha was .79 and .80 at Time 1 and Time 2, respectively, with a test-retest reliability of r = .63.

KeyMath3 3. KeyMath3 (Connolly, 2007) is a norm-referenced measure designed to assess mathematics ability from ages 4 years, 6 months to 21 years. It is often used to identify gaps in students’ knowledge and skills for the purpose of tailoring individualized interventions. It comprises three broad content areas aligned with the National Council of Teachers of Mathematics (NCTM)’s Principles and Standards for School Mathematics (2000). Three subtests from the Basic Concepts area were used: Numeration, Geometry, and Measurement. Within each subtest, administration was discontinued if and when a child incorrectly answered four consecutive items incorrectly. The Numeration subtest (up to 49 items possible) covers a wide range of foundational mathematical concepts and skills from basic number concepts through place value and eventually exponents and square roots. Children were administered at most 17 Numeration items in the fall, with internal reliability of those items at that time point being .75; they were administered at most 25 Numeration items in the spring, with an internal reliability of those items at that time point being .80. Test-retest reliability of the Numeration subtest was .67.

The Geometry subtest (up to 36 items possible) assesses students' ability to identify and analyze two- and three-dimensional shapes in addition to aspects of spatial relationships and reasoning. Children were administered at most 25 Geometry items in the fall, with internal reliability of those items at that time point being .75; they were administered at most 29 Geometry items in the spring, with an internal reliability of those items at that time point being .74. Test-retest reliability of the Geometry subtest was .54.

The Measurement subtest (up to 40 items possible) covers the ability to compare and measure objects in standard and non-standard units across a range of domains, including distance, size, volume, money, and time. Children were administered at most 12 Measurement items in the fall, with internal reliability of those items at that time point being .73; they were administered at most 21 Measurement items in the spring, with an internal reliability of those items at that time point being .81. Test-retest reliability of the Measurement subtest was .67. Raw scores from these three subtests were summed to create aKeyMath3 Composite score, which in this sample had an internal reliability of .87 and .90 in the fall and spring, respectively, and a test-retest reliability of r = .79.

Test of Early Mathematics Ability. Unlike Woodcock-Johnson Applied Problems and KeyMath3, which were designed to test math into adulthood, the TEMA (Ginsburg & Baroody, 2003) is a 72-item assessment specifically designed for young children from three to eight years old. It also includes items that assess both informal math concepts and skills (those learned in the home or neighborhood) and formal math concepts and skills (those learned in school). Informal concepts and skills include, among others, small number perception, modeling addition and subtraction using objects, understanding part-whole relations, verbal counting, understanding cardinality, and constructing a mental number line. Formal concepts and skills include reading and writing numerals, addition and subtraction facts, and written addition and subtraction procedures. Starting point for administration of TEMA items was adjusted based on a child's age and assessment administration was discontinued if a child answered 8 consecutive items incorrectly. In this sample, at least one child was assumed to have answered all 72 items at each time point, and the internal reliability was 0.97 and .65 at Time 1 and Time 2, respectively, and test-retest reliability was r = .79.

Analytic Plan

Analyses were conducted in Mplus v. 7.0 (Muthen & Muthen, 2007) and using maximum likelihood estimation, robust standard errors to account for children being nested within classrooms throughout the school year, and multiple imputation in order to address missing data.

Multiple regressions. Multiple regression models were used to examine associations between cognitive skill predictors and mathematics performance at Time 1 (Models 1A through 1U) and at Time 2, controlling for Time1 (Models 2A through 2U). Analyses were conducted using composite mathematics (average of z-scores across the three mathematics measures) and, as follow-up analyses, each measure of mathematics performance – including models for each of the KeyMath3 subtests. For each time point, a model with covariates only (e.g., as in Models 1A and 2A) was tested prior to simultaneous entry of EF and VMI (e.g., as in Models 1B and 2B) to understand how much variance these skills explained above and beyond covariates. All models included gender (0 = female; 1 = male), grade (0 = kindergarten; 1 = first grade), and treatment condition (0 = not offered treatment; 1 = offered treatment) as covariates. Age was not included as a covariate because it was not a significant predictor of mathematics performance after controlling for grade. Further, the models predicting Time 2 mathematics performance included Time 1 mathematics performance as a predictor. To test potential interaction effects between EF and VMI, interaction terms were added to the models (e.g., Models 1C and 2C). EF and VMI were centered prior to creating the interaction variable in order to ease interpretation of results. For each regression, a Wald test was run in order to determine whether adding EF and VMI (or their interaction) as predictors increased the amount of explained variance in the mathematics performance outcomes. Then, the regression coefficients were examined to understand whether and to what extent each skill (or interaction) predicted mathematics performance above and beyond other predictors, as indicated by the statistical significance and magnitude of each coefficient, respectively.

Missing data. Table 1 makes evident that missing data was a minor issue in this study. Most variables ranged from 0 to 14 missing observations in the amount of missing data. NEPSY Attention/EF domain scores could not be calculated for 13 children due to experimenter error, but other missing data was due to study attrition. As mentioned above, analyses used MI, conducted in Mplus. MI in Mplus uses Markov Chain Monte Carlo simulations, using 100 iterations to generate each imputed data set (see Asparouhov & Muthen, 2010). We imputed 10 data sets, on which analyses were conducted and pooled.

Observational studies with experimental data: Sensitivity analyses. Because these analyses were based on data collected as part of an intervention study, it is reasonable to be concerned about whether the intervention may affect relations among constructs. In this particular study, the treatment group's condition was designed to improve fine motor skills, but was also found to improve EF and visuospatial skills for the first of the two cohorts represented in the current study sample (Grissmer et al., 2013). Thus, sensitivity analyses were conducted to understand whether and to what extent results were affected by the intervention. Specifically, the same analyses as described above were conducted, but with models including interaction term(s) between variables of interest (EF, VMI, and their interaction) and a dummy variable coded as 1 for children from the first cohort who were assigned to treatment and as 0 for all other children. These terms were included in imputation models as well as analytic models. We conducted Wald tests in order to understand whether adding these interaction terms increased the amount of variance explained in the outcome in question. We also examined the statistical significance of the coefficients for these terms in analytic models, which would indicate whether associations between constructs was different for the first cohort's treatment group, as compared to all other children.

[TABLE 1 – MAY NEED OWN LANDSCAPE ORIENTED PAGE]

Results

Children in this sample performed low on EF skills relative to the normative NEPSY sample (See Table 1): average percentile ranks for the Attention/EF domain were in the 17th percentile for kindergarteners and 25th percentile for first graders. These descriptive statistics demonstrate the extent to which the current, low-SES sample would be considered less advantaged than those of other studies discussed above (e.g., Blair & Razza, 2007; Cameron et al., 2012).

[TABLE 2 - LANDSCAPE]

Bivariate correlations (see Table 2) were computed among all EF and VMI assessments and mathematics outcomes at both time points. In general, all correlations were in the expected direction. For instance, EF domain scores and VMI scores were positively correlated (r = .33, p < .001), as were EF and VMI with each mathematics assessment.

Contributions of Time 1 EF and VMI to Mathematics Performance

Regression model estimates for both concurrent and longitudinal analyses for composite mathematics scores are shown in Table 3. Regression model estimates for individual mathematics measure scores for concurrent and longitudinal analyses can be found in Supplemental Tables S4 and S5, respectively.

[TABLE 3]

Adding EF and VMI to the concurrent models (e.g., comparing Model 1A and 1B) significantly increased the amount of variance explained, above that explained by covariates, in Time 1 composite mathematics performance by 16% (Wald statistic = 28.52, Δdf = 2, p < .001; see Table 3). The increase in amount of variance explained was also significant for all three mathematics assessments, as well as each KeyMath3 subtest: by at least 10% (for KeyMath3 Geometry; Wald statistic = 14.9, Δdf = 2, p < .001) and at most 16% (for the KeyMath3 Composite, Wald statistic = 29.95, Δdf = 2, p < .001; see Table S4). Both skills had significant associations with Time 1 composite mathematics performance as well as across all individual measures. The magnitude of the association between EF and composite mathematics was β = .28 (p < .001). For each individual mathematics measure, the coefficient for EF was at least β = .16 (KeyMath3 Geometry, p < .05) and up to β = .31 (KeyMath3 Composite and KeyMath3 Measurement, p < .001). The magnitude of the association between VMI and composite mathematics was β = .26 (p < .001). For each individual mathematics measure, the coefficient for VMI was at least β = .20 (KeyMath3 Measurement, p < .01) and up to β = .27 (KeyMath3 Geometry and KeyMath3 Numeracy, p < .001).

Adding EF and VMI to the models predicting Time 2 (e.g., comparing Model 2A and 2B) performance significantly increased the amount of variance explained in composite mathematics by 3% (Wald Statistic = 16.82, $\Delta df = 2$, $p < .001$; see Table 3). The increase in amount of variance explained was also significant for all three mathematics assessments, as well as each KeyMath3 subtest: by at least 3% (KeyMath3 Composite, Wald Statistic = 69.77, $\Delta df = 2$, $p < .001$) and up to 7% (KeyMath3 Geometry, Wald Statistic = 85.95, $\Delta df = 2$, $p < .001$; see Table S5). Further, both skills significantly predicted improvement in composite mathematics performance and across almost all individual mathematics measures. The partial effect of EF was significant for composite mathematics ($\beta = .11$, $p < .01$), as well as for five out of six individual measures, and was as large as $\beta = .20$ (KeyMath3 Measurement, $p < .01$). The exception to this statement was KeyMath3 Geometry ($\beta = .10$, $p = .11$). The partial effect of VMI was significant for composite mathematics ($\beta = .15$, $p < .01$), as well as for all six measures, ranging from $\beta = .13$ (KeyMath3 Measurement, $p < .05$) to $\beta = .26$ (KeyMath3 Geometry, $p < .001$).

Potential Interaction Effects between EF and VMI

To test the compensatory pattern of cognitive skills on mathematics performance, we added an interaction term between EF and VMI to the models described immediately above (e.g., Models 1C and 2C). The interaction term did not explain significant variance above and beyond covariates and EF and VMI for composite mathematics performance either concurrently (Wald statistic = 0.00, $\Delta df = 2$, $p = .96$) or longitudinally (Wald statistic = 0.54, $\Delta df = 2$, $p = .46$; see Table 3). The change in explained variance was statistically significantly different from zero for only two individual outcomes: concurrent KeyMath3 Geometry (Table S4, Model 1O: Wald statistic = 5.08, $\Delta df = 1$, $p < .05$; $\beta = -.17$, $p < .05$) and Time 2 TEMA (Table S5, Model 2I: Wald statistic = 2.06, $\Delta df = 1$, $p < .05$; $\beta = -.16$, $p < .05$). The nature of the interaction for both of these outcomes was compensatory, as hypothesized: children with *either* high VMI *or* high EF performed better on Time 1 KeyMath3 Geometry and Time 2 TEMA than children who were weak in both skills (Figures 1 and 2 illustrating these relations are available as part of online, supplementary materials). We note that while the nature of this interaction for these outcomes was as hypothesized, we did not hypothesize that this interaction would exist for only two of 16 tests for said interaction.

Covariate Effects

In the concurrent models, grade was the only covariate related to mathematics outcomes after controlling for EF and VMI: not surprisingly, first graders outperformed kindergarteners in composite mathematics and across all individual measures. There was a gender effect in two of the covariates-only concurrent models, with girls outperforming boys on WJ Applied Problems and KeyMath3 Geometry, but this advantage did not persist after inclusion of EF and VMI. Across longitudinal models, grade was, again, the only covariate related to outcomes after controlling for EF and VMI: first graders outperformed kindergartners across all measures, even after controlling for Time 1 performance. In other words, the longitudinal models suggest first graders showed more improvement in mathematics performance, on average, than kindergarteners over the course of the school year.

Sensitivity Analyses: Testing for Interference of Intervention in Producing Results

In order to ensure that intervention effects from the larger study did not interfere with relations observed among constructs at Time 2 (the post-intervention time point), we ran analyses identical to those described above, but including an interaction term between variables of interest (i.e., EF, VMI, and their interaction) and a dummy variable indicating whether each child was in the treatment group in the study cohort which had significant treatment effects. In no model did this interaction term add significant variance to the outcome. For Time 2 TEMA, despite an insignificant Wald statistic, the treatment X EF coefficient was significant ($\beta = -.12$, $p < .05$), but this indicates that the association between EF and Time 2 TEMA was weaker for the group of children offered treatment. Therefore, if being offered an opportunity to participate in the intervention affected our results with respect to this outcome, it did so by decreasing the magnitude of the partial EF effect. Thus, we conclude that any intervention effects are not responsible for the patterns of associations we have found and as described above; if anything, intervention effects may have attenuated observed associations.

Discussion

This study examined the extent to which two cognitive skills, EF and VMI, measured at the beginning of the school year (Time 1), were associated with concurrent mathematics performance and longitudinally with improvement in mathematics performance through the end of the school year (Time 2) in a low-SES sample of kindergarteners and first graders. Results suggest that both skills are associated with concurrent performance as well as improvement in performance on overall mathematics performance, as well as on most individual mathematics assessments included in this study. Results do not support a compensatory pattern, or any other type of interaction effect, for these cognitive skills in relating to overall mathematics performance, nor to most (12 out of 14) individual measures of mathematics performance, in early elementary school, except for concurrent KeyMath3 geometry performance and improvement in TEMA performance.

This study replicated previously established links between EF and early academic skills before, during, and after the transition to formal schooling (e.g. Blair & Razza, 2007; Duncan et al., 2007; McClelland et al., 2007). In particular, these results are closely aligned with those reported by Blair and Razza (2007) which found, in a low-income sample, that preschoolers with higher EF had stronger early math and reading skills at the end of kindergarten. The current study finds much the same, with the added strength of analyses controlling for Time 1 mathematics performance; in other words, we found that children who began the school year with higher EF made larger improvements in mathematics performance than those who entered with lower EF. This was the case for overall mathematics, as well as for five out of six mathematics outcomes, the exception being KeyMath3 Geometry.

Previous work has underscored the importance of EF for successfully meeting the new challenges of formal schooling, such as more complex academic tasks that require strategy selection and manipulation of mental models (Bull & Scerif, 2001; Fuchs et al., 2010b) and navigating the classroom environment (Rimm-Kaufman, Pianta, & Cox, 2000). Children with low EF are likely to have difficulty with both, which would explain why EF predicts Time 2 mathematics performance on most assessments even after taking Time 1 mathematics performance into account.

This study extended past research by suggesting that VMI relates to both concurrent mathematics performance as well as improvement in mathematics performance, even after controlling for EF. Evidence for a link between visuospatial skills and mathematics ability spans brain imaging (e.g. Hubbard, Piazza, Pinel, & Dehaene, 2005), disabilities (e.g. O'Hearn & Landau, 2007), and adult numerical cognition research (e.g. Dehaene, Bossini, & Giraux, 1993). To date, however, relatively fewer studies have addressed visuospatial skills and mathematics performance in young children, especially when controlling for other potential confounds, such as EF. This study establishes that even if children have strong executive skills, VMI appears to have an additional, unique association with improvement in most measures of mathematics performance. This could be due to the heavy incorporation of manipulatives and other visual and tactile aids in early elementary mathematics instruction (Guarino, Dieterle, Bargagliotti, & Mason, 2013). Even if children can pay attention and control impulses, they also need to be able to visualize and work with materials. Alternatively, the pattern of findings could be due to co-development in visuospatial and mathematics skills, which rely on similar neural networks (Hubbard et al., 2005).

That both EF and VMI were related to concurrent KeyMath3 Geometry, but only VMI predicted improvement in KeyMath3 Geometry, could reflect this particular assessment's emphasis on spatial reasoning. For example, in this assessment students are asked to describe the location of a person in a picture, match pictures of a kitten in different positions, identify left and right, and identify what a picture would look like if rotated. These tasks are less formal than items in most of the other assessments in this study, and therefore the lack of longitudinal EF effect may reflect that these skills are less explicitly taught in instruction throughout the year.

In slight contrast to other studies suggesting a compensatory effect between EF and VMI for supporting emergent literacy and classroom behavior among 3- to 5-year-olds (Cameron et al., 2015), this study found no interaction effect between EF and VMI on overall mathematics and on most aspects of mathematics performance, except for on concurrent geometry performance and on improvement on TEMA. This discrepancy between this study and Cameron and colleagues (2015) could be due to age differences where cognitive skills are more likely to compensate for each other in younger children. Consistent with that line of reasoning, Cameron et al. (2012) tested but did not find that EF and VMI interacted in relation to kindergarten academic achievement, including in mathematics problem solving. This study largely suggests the same, but demonstrates that such an interaction may exist for other types of mathematics performance (i.e., concurrent Geometry and improvement in informal mathematics, as measured by TEMA).

The nature of the interaction between EF and visuospatial skills for beginning-of-year (concurrent) geometry performance and end-of-year (later) TEMA performance was such that children performed almost equally well if they had high EF, high VMI, or both, but not if they scored low on both. As mentioned previously, the KeyMath3 Geometry subtest included many informal items; likewise, a majority of items answered correctly by children on the TEMA at both the beginning and end of the year measured informal math skills. Thus, this interaction may suggest that strong visuospatial skills may compensate for low EF in using or developing informal mathematics skills. For example, a child who has difficulty planning out or following the steps of a simple addition problem may nevertheless arrive at the correct answer if he or she can manipulate and arrange objects in space to model the problem. The presence of an interaction at the end, but not the beginning, of the year for TEMA could suggest that children with low EF may rely on their relatively stronger VMI to benefit from classroom instruction, to the extent such instruction is provided.

Strengths and Limitations

This study was one of the first to utilize normed neuropsychological assessments of both EF and visuospatial skills to examine their contributions to early mathematics performance. While the small sample size allowed for thorough testing, it also limited the generalizability of the results. These low-SES children from one school district are not likely to be representative of low-SES early elementary-aged children in all regions of the country.

Another limitation is the observational nature of the study. While controlling for prior mathematics performance in our longitudinal analyses lends some support for the notion that EF and visuospatial skills might contribute to mathematics learning, our analyses do not demonstrate causal relationships. On a related note, we have not controlled for several potential confounds in our analyses, such as language and intelligence; whether our results would hold after doing so cannot be known.

Measurement of mathematics performance here was more thorough than in previous studies using only one measure, in that three psychometrically validated measures of mathematics were used. This minimizes the possibility that the present findings of unique associations between EF and mathematics and VMI and mathematics are unique to a certain type of mathematics measure or set of skills, as they were consistent across most measures. On the other hand, we acknowledge that running models across this many mathematics assessments effectively resulted in 16 separate tests seeking an EF X VMI interaction, and therefore, we may have found significant interactions for Time 1 Geometry and Time 2 TEMA by chance.

Measurement of cognitive skills, on the other hand, may be a concern despite the use of validated assessments. First, reliabilities for several EF subtests were somewhat low, and may reflect the fact that the NEPSY assessment was normed/standardized on a nationally representative sample of children, whereas the sample included here is comprised largely of children from low-SES families. This low reliability suggests larger than desired measurement error and, ultimately, reduced power for detecting effects. Thus, imprecision in EF measurement could be to blame for why EF did not significantly predict improvement in KeyMath3 Geometry scores, despite having a positive regression coefficient. This underscores the need for psychometric work in low-SES populations, who are traditionally under-represented in psychological research. Second, the assessments used for EF simultaneously tapped multiple components of EF, disallowing independent tests of associations between various aspects of EF and mathematics. Given that different components of EF have been differentially related to academic outcomes in similarly-aged populations (Bull, Epsy, & Wiebe, 2008), future research should consider whether such differential relations hold even in the context of controlling for VMI.

Finally, given that the data utilized in this study were ultimately collected for a purpose other than examining the relation between EF, VMI, and mathematics performance (i.e., were collected as part of an evaluation study), the analyses should be regarded as exploratory. Future replication studies should be conducted in order to confirm whether the pattern of results holds.

Implications and Conclusion

These analyses may inform thinking about the connection between cognitive development and academic achievement, including how the former might support the latter. This consideration may be especially important for children who come from low-SES households with fewer opportunities and greater challenges in their development of certain cognitive skills not typically explicitly taught or trained in school. In this study, EF and VMI were robust and unique predictors of improvement in mathematics performance in a sample of low-SES kindergartners and first graders. These findings are consistent with previous research and with the argument for the inclusion of these cognitive skills in the focus of early education policies and curricula (Blair, 2002, 2003; Rimm-Kaufman & Pianta, 2000; Ursache et al., 2012; Uttal & Cohen, 2012).

Importantly, both EF and VMI are sensitive to intervention and experience (Diamond, Barnett, Thomas, & Munro, 2007; Newcombe et al., 2013; Tang & Posner, 2009; Uttal et al., 2012). Promising intervention approaches targeting EF include novel, cohesive curricula implementable in preschool and kindergarten classrooms (Bierman & Torres, in press; Diamond, 2012). There have been comparatively fewer implementations of curricula explicitly fostering visuospatial abilities, such as VMI, in early childhood, which has been called “the orphan of the academic curriculum” (p. 44, Newcombe, et al., 2013). Some researchers have begun efforts to flesh out the “When, Why, and How” of incorporating visuospatial development in early education efforts, proposing that explicit training of visuospatial skills has the potential to improve learning in mathematics and other science disciplines (e.g., Uttal & Cohen, 2012). For example, simple activities such as the use of spatial language, maps or models, and sensorimotor materials appear to support visuospatial development in young children and may be easily implementable within current early childhood curricula (see Newcombe et al., 2013). Thus, future research should investigate the extent to which the associations found here represent causal relationships and, relatedly, whether fostering these abilities transfers to improved mathematics learning later on.

The results of this study lend tentative support to those who advocate for a more diverse curriculum, with a broader focus on child development and incorporation of activities not explicitly tied to academic outcomes (e.g. Elkind, 2012; Stipek, 2006). Specifically, EF and visuospatial skills may be two cognitive skills that are theoretically important for children’s early success in mathematics in elementary school. Given the demographics of our sample, our results suggest this is the case for children of color from low-income families in addition to previously studied, more socio-demographically advantaged families. Further research is required to evaluate the promotion of these skills as a means to improve academic outcomes for children, and such efforts are already underway (e.g., Grissmer et al., 2013). Results of such efforts should provide insights for how best to holistically support academic success for all children.

References

Akshoomoff, N. A., & Stiles, J. (1995). Developmental trends in visuospatial analysis and planning: I. Copying a complex figure. *Neuropsychology, 9*(3), 364–377. doi:10.1037/0894-4105.9.3.364

Assel, M. A., Landry, S. H., Swank, P., Smith, K. E., & Steelman, L. M. (2003). Precursors to Mathematical Skills: Examining the Roles of Visual-Spatial Skills, Executive Processes, and Parenting Factors. *Applied Developmental Science, 7*(1), 27–38. doi:10.1207/S1532480XADS0701_3

Bassok, D., Latham, S., & Rorem, A. (2015). Is Kindergarten the New First Grade?. Working Paper - EdPolicy Works, University of Virginia, Charlottesville, VA.

Becker, D. R., Miao, A., Duncan, R., & McClelland, M. M. (2014). Behavioral self-regulation and executive function both predict visuomotor skills and early academic achievement. Early Childhood Research Quarterly, 29(4), 411-424. doi:10.1016/j.ecresq.2014.04.014

Beery, K. E., & Buktenica, N. A. (1997). Beery-Buktenica Developmental Test of Visual-Motor Integration (4th ed.). Modern Curriculum Press.

Bierman, K. L. & Torres, M. (in press). Promoting the Development of Executive Functions through Early Education and Prevention Programs. J. A. Griffin, L.S. Freund, & P. McCardle (Eds.) (in press). *Executive function in preschool age children: Integrating measurement, neurodevelopment and translational research.* Washington, D.C.: American Psychological Association.

Blair, C. (2002). School readiness: Integrating cognition and emotion in a neurobiological conceptualization of children’s functioning at school entry. *American Psychologist, 57*(2), 111–127. doi:10.1037/0003-066X.57.2.111

Blair, C. (2003). Behavioral inhibition and behavioral activation in young children: Relations with self-regulation and adaptation to preschool in children attending Head Start. *Developmental Psychobiology, 42*(3), 301–311.

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development, 78*(2), 647–63. doi:10.1111/j.1467-8624.2007.01019.x

Bongard, S., & Nieder, A. (2010). Basic mathematical rules are encoded by primate prefrontal cortex neurons. *Proceedings of the National Academy of Sciences*, *107*(5), 2277-2282.

Bradley, R. H., & Caldwell, B. M. (2013). 174 Children: A study of the relationship between home environment and cognitive development during the first 5 years. In Gottfried, A. W. (Ed.). (2013). *Home environment and early cognitive development: Longitudinal research*. Academic Press.

Bull, R., & Scerif, G. (2001). Executive functioning as a predictor of children's mathematics ability: inhibition, switching, and working memory. *Developmental Neuropsychology, 19*(3), 273–93. doi:10.1207/S15326942DN1903_3

Cameron, C. E., Brock, L. L., Hatfield, B. H., Cottone, E. A., Rubinstein, E., LoCasale-Crouch, J., & Grissmer, D. W. (2015). Visuomotor integration and inhibitory control compensate for each other in school readiness. *Developmental Psychology, 51*(11), 1529-1543. doi:10.1037/a0039740

Cameron, C. E., Brock, L. L., Murrah, W. M., Bell, L. H., Worzalla, S. L., Grissmer, D., & Morrison, F. J. (2012). Fine motor skills and executive function both contribute to kindergarten achievement. *Child Development, 83*(4), 1229–44. doi:10.1111/j.1467-8624.2012.01768.x

Cameron, C. E., Cottone, E. A., Murrah, W. M., & Grissmer, D. W. (2016). How Are Motor Skills Linked to Children's School Performance and Academic Achievement?. *Child Development Perspectives 10*(2).

Cameron Ponitz, C., M. M. McClelland, et al. (2009). A structured observation of behavioral self-regulation and its contribution to kindergarten outcomes. *Developmental Psychology 45*(3): 605-619.

Carlson, A. G., Rowe, E., & Curby, T. W. (2013). Disentangling fine motor skills' relations to academic achievement: The relative contributions of visual-spatial integration and visual-motor coordination. *The Journal of genetic psychology*, *174*(5), 514-533.

Carlton, M. P., & Winsler, A. (1999). School readiness: The need for a paradigm shift. School *Psychology Review, 28(3)*, 338–352.

Clark, R., Kirschner, P. A., & Sweller, J. (2012). Putting students on the path to learning: The case for fully guided instruction. *American Educator 36*(1), 6-11.

Clark, C. A., Sheffield, T. D., Wiebe, S. A., & Espy, K. A. (2013). Longitudinal associations between executive control and developing mathematical competence in preschool boys and girls. *Child Development 84*(2), 662-677.

Connolly, A. J. (2007). KeyMath3TM Diagnostic Assessment. Minneapolis, MN: Pearson.

Cragg, L., & Gilmore, C. (2014). Skills underlying mathematics: The role of executive function in the development of mathematics proficiency. *Trends in Neuroscience and Education*, *3*(2), 63-68.

Culbertson, W. C., & Zillmer, E. A. (1998). The Tower of London DX: A standardized approach to assessing executive functioning in children. *Archives of Clinical Neuropsychology, 13*(3), 285-301.

Decker, S. L., Englund, J. A., Carboni, J. A., & Brooks, J. H. (2011). Cognitive and developmental influences in visual-motor integration skills in young children. Psychological Assessment, 23, 1010-1016. http://dx.doi.org/10.1037/a0024079

Dehaene, S. (2011). *The Number Sense: How the Mind Creates Mathematics* (Rev Upd., p. 352). New York: Oxfrod Univeristy Press.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General, 122*(3), 371–396. doi:10.1037/0096-3445.122.3.371

Dehaene, S., & Cohen, L. (2007). Cultural recycling of cortical maps. Neuron, 56(2), 384-398.

Dehaene, S., & Mehler, J. (1992). Cross-linguistic regularities in the frequency of number words. *Cognition*, *43*(1), 1–29. doi:http://dx.doi.org/10.1016/0010-0277(92)90030-L

Diamond, A. (2006). The early development of executive functions. In E. Bialystok & F. I. M. Craik (Eds.), *Lifespan Cognition: Mechanisms of Change* (pp. 70–95). NYC, NY: Oxfrod Univeristy Press.

Diamond, A. (2012). Activities and Programs That Improve Children’s Executive Functions. *Current Directions in Psychological Science*, *21*(5), 335–341. doi:10.1177/0963721412453722

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135–68. doi:10.1146/annurev-psych-113011-143750

Diamond, A., Barnett, S. W., Thomas, J., & Munro, S. (2007). Preschool program improves cognitive control. *Science, 318*(5855), 1387–8. doi:10.1126/science.1151148

Dilworth-Bart, J. E. (2012). Does executive function mediate SES and home quality associations with academic readiness? *Early Childhood Research Quarterly, 27*(3), 416-425.

Doricchi, F., Guariglia, P., Gasparini, M., & Tomaiuolo, F. (2005). Dissociation between physical and mental number line bisection in right hemisphere brain damage. *Nature Neuroscience, 8*(12), 1663–5. doi:10.1038/nn1563

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., ... & Sexton, H. (2007). School readiness and later achievement. *Developmental psychology*, *43*(6), 1428.

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling, 8*(3), 430-457.

Floyer-Lea, A., & Matthews, P. M. (2004). Changing brain networks for visuomotor control with increased movement automaticity. *Journal of neurophysiology, 92*(4), 2405-2412. doi:10.1152/jn.01092.2003

Fuchs, L. S., Geary, D. C., Compton, D. L., Fuchs, D., Hamlett, C. L., & Bryant, J. D. (2010). The contributions of numerosity and domain-general abilities to school readiness. *Child Development, 81*(5), 1520–33. doi:10.1111/j.1467-8624.2010.01489.x

Fuchs, L. S., Geary, D. C., Compton, D. L., Fuchs, D., Hamlett, C. L., Seethaler, P. M., Bryant, J. D., et al. (2010). Do different types of school mathematics development depend on different constellations of numerical versus general cognitive abilities? *Developmental Psychology*, *46*(6), 1731–46. doi:10.1037/a0020662

Geary, D. C. (2011). Cognitive predictors of achievement growth in mathematics: a 5-year longitudinal study. *Developmental Psychology, 47*(6), 1539.

Ginsburg, H. P., & Baroody, A. J. (2003). *Test of Early Mathematics Ability, Third Edition (TEMA-3)*. Austin, TX: pro-ed.

Grissmer, D., Mashburn, A., Cottone, E., Brock, L., Murrah, W., Blodgett, J., & Cameron, C. (2013). The efficacy of minds in motion on children's development of executive function, visuo-spatial and math skills. Society for Research on Educational Effectiveness (SREE), Washington, DC.

Grissmer, D., Grimm, K. J., Aiyer, S. M., Murrah, W. M., & Steele, J. S. (2010). Fine motor skills and early comprehension of the world: two new school readiness indicators. *Developmental Psychology, 46*(5), 1008–17. doi:10.1037/a0020104

Guarino, C., Dieterle, S. G., Bargagliotti, A. E., & Mason, W. M. (2013). What can we learn about effective early mathematics teaching? A framework for estimating causal effects using longitudinal survey data. *Journal of Research on Educational Effectiveness, 6*(2), 164-198. doi:10.1080/19345747.2012.706695

Gunderson, E. A., Ramirez, G., Beilock, S. L., & Levine, S. C. (2012). The relation between spatial skill and early number knowledge: the role of the linear number line. *Developmental Psychology, 48*(5), 1229–41. doi:10.1037/a0027433

Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. National Bureau of Economic Research Working Paper Series, No. 12006.

Hirsh-Pasek, K. (2009). A mandate for playful learning in preschool: presenting the evidence. Oxford University Press.

Hubbard, E. M., Piazza, M., Pinel, P., & Dehaene, S. (2005). Interactions between number and space in parietal cortex. Nature Reviews. *Neuroscience, 6*(6), 435–48. doi:10.1038/nrn1684

Jacques, S., & Marcovitch, S. (2010). Development of executive function across the life span. In W. F. Overton & R. M. Lerner (Eds.), *The handbook of life-span development* (pp. 431–466). Hoboken, NJ, US: John Wiley & Sons Inc.

Kim, H., Byers, A. I., Cameron, C. E., Brock, L. L., Cottone, E. A., & Grissmer, D. W. (2016). Unique contributions of attentional control and visuomotor integration on concurrent teacher-reported classroom functioning in early elementary students. *Early Childhood Research Quarterly, 36*, 379-390.

Korkman, M., Kirk, U., & Kemp, S. L. (1998). *NEPSY: A Developmental Neuropsychological Assessment*. Psychological Corp.

Korkman, M., Kirk, U., & Kemp, S. L. (2007). *NEPSY-II* (p. 290). Bloomington, MN: Pearson.

LeFevre, J. A., Berrigan, L., Vendetti, C., Kamawar, D., Bisanz, J., Skwarchuk, S. L., & Smith-Chant, B. L. (2013). The role of executive attention in the acquisition of mathematical skills for children in Grades 2 through 4. *Journal of Experimental Child Psychology, 114*(2), 243-261.

Luo, Z., Jose, P. E., Huntsinger, C. S., & Pigott, T. D. (2007). Fine motor skills and mathematics achievement in East Asian American and European American kindergartners and first graders. *British Journal of Developmental Psychology, 25*(4), 595–614. doi:10.1348/026151007X185329

McClelland, M. M., Acock, A. C., Piccinin, A., Rhea, S. A., & Stallings, M. C. (2013). Relations between Preschool Attention Span-Persistence and Age 25 Educational Outcomes. *Early Childhood Research Quarterly, 28*(2), 314–324. doi:10.1016/j.ecresq.2012.07.008

McClelland, M. M., Cameron, C. E., Connor, C. M., Farris, C. L., Jewkes, A. M., Morrison, F. J., Mashburn, A. J., et al. (2007). Links between behavioral regulation and preschoolers' literacy, vocabulary, and math skills. *Developmental Psychology, 43*(4), 947–59. doi:10.1037/0012-1649.43.4.947

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(1), 49–100. doi:http://dx.doi.org/10.1006/cogp.1999.0734

National Council of Teachers of Mathmatics. (2000). *Principles and Standards for School Mathematics*. Reston, VA: Author.

Newcombe, N. S., Uttal, D. H., & Sauter, M. (2013). Spatial Development. In P. D. Zelazo (Ed.), *Oxford Handbook of Developmental Psychology*. Oxford University Press.

O'Hearn, K., & Landau, B. (2007). Mathematical skill in individuals with Williams syndrome: evidence from a standardized mathematics battery. *Brain and Cognition*, *64*(3), 238–46. doi:10.1016/j.bandc.2007.03.005

Potter, D., Mashburn, A. J., & Grissmer, D. (2013). The family, neuroscience, and academic skills: An interdisciplinary account of social class gaps in children's test scores. *Social Science Research*, *42*(2), 446–64. doi:10.1016/j.ssresearch.2012.09.009

Rimm-Kaufman, S. E., & Pianta, R. C. (2000). An Ecological Perspective on the Transition to Kindergarten: A Theoretical Framework to Guide Empirical Research. *Journal of Applied Developmental Psychology, 21*(5), 491–511.

Rimm-Kaufman, S. E., Pianta, R. C., & Cox, M. J. (2000). Teachers' judgments of problems in the transition to kindergarten. *Early Childhood Research Quarterly*, *15*(2), 147-166.

Romine, C. B., & Reynolds, C. R. (2005). A model of the development of frontal lobe functioning: Findings from a meta-analysis. *Applied Neuropsychology, 12*(4), 190–201. doi:10.1207/s15324826an1204_2

Shepard, R. N., & Metzler, J. (1971). Mental Rotation of Three-Dimensional Objects. *Science, 171*(3972), 701–703. doi:10.2307/1731476

Sortor, J. M., & Kulp, M. T. (2003). Are the results of the Beery-Buktenica Develomental Test of Visual-Motor Integration and its subtests related to achievement test scores? *Optometry & Vision Science*, *80*, 758-763. doi:1040-5488/03/8011-0758/0

Stevenson, H. W., & Newman, R. S. (1986). Long-term prediction of achievement and attitudes in mathematics and reading. *Child Development, 57*(3), 646–59.

Stiles, J., Paul, B., & Ark, W. (2008). The development of visuospatial processing. In C. A. N. M. Luciana (Ed.), *Handbook of Developmental Cognitive Neuroscience (2nd ed.)* (pp. 521–540). Cambridge, MA, US: MIT Press.

Stipek, D. J. (2006). No Child Left Behind Comes to Preschool. *The Elementary School Journal, 106*(5), 455–466.

Ursache, A., Blair, C., & Raver, C. C. (2012). The Promotion of Self-Regulation as a Means of Enhancing School Readiness and Early Achievement in Children at Risk for School Failure. *Child Development Perspectives, 6*(2), 122–128. doi:10.1111/j.1750-8606.2011.00209.x

Uttal, D. H., & Cohen, C. A. (2012). Spatial thinking and STEM education: When, why and how. *Psychology of Learning and Motivation, 57*(2), 147-181.

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013). The malleability of spatial skills: a meta-analysis of training studies. *Psychological Bulletin, 139*(2), 352

Verdine, B. N., Golinkoff, R. M., Hirsh-Pasek, K., Newcombe, N. S., Filipowicz, A. T., & Chang, A. (2014). Deconstructing building blocks: Preschoolers' spatial assembly performance relates to early mathematical skills. *Child Development*, *85*(3), 1062-1076.

Verdine, B. N., Irwin, C. M., Golinkoff, R. M., & Hirsh-Pasek, K. (2014). Contributions of executive function and spatial skills to preschool mathematics achievement. *Journal of Experimental Child Psychology*, *126*, 37-51.

Welsh, M. C., Pennington, B. F., & Groisser, D. B. (1991). A normative-developmental study of executive function: A window on prefrontal function in children. *Developmental Neuropsychology*, *7*(2), 131-149.

Willoughby, M. T., Blair, C. B., Wirth, R. J., & Greenberg, M. (2012). The measurement of executive function at age 5: psychometric properties and relationship to academic achievement. *Psychological Assessment, 24*(1), 226.

Woodcock, R. W., McGrew, K. S., Mather, N., & Schrank, F. (2001). *Woodcock-Johnson III Tests of Academic Achievement (3rd ed.)*. Rolling Meadows, IL: Riverside Publishing Company.

Tables

Table 1. Sample Characteristics and Descriptives

	Kindergarteners					First Graders				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
Age T1 (months)	88	66.55	3.96	60	75	70	79.03	4.53	69	92
Age T2 (months)	84	73.2	3.96	67	83	68	85.42	4.48	79	99
Gender (1 = Male)	89	0.52	0.5	0	1	73	0.48	0.5	0	1
Race (1 = Black)	89	0.94	0.23	0	1	73	0.9	0.3	0	1
FRL (1 = eligible)	83	0.96	0.19	0	1	71	0.93	0.26	0	1
Offered Treatment (1 = Offered Treatment)	89	0.51	0.5	0	1	73	0.58	0.5	0	1
EF Core Domain Score T1	75	81.35	13.2	53	121	67	85.57	15.1	54	121
EF Core Domain Percentile T1	75	17.27	18.8	0.1	92	67	25.02	25.6	0.1	92
Design Copy (VMI) Score - T1	87	-3.49	7.66	-30	12	70	-0.05	1.15	-2.3	2.6
WJ - Applied Problems Raw Score - T1	89	12.66	3.58	2	20	73	17.91	3.95	7	25
WJ - Applied Problems Raw Score - T2	84	17.46	3.73	5	26	68	23.06	4.03	12	32
TEMA Raw Score - T1	89	14.70	8.10	0	32	72	28.99	9.12	4	57
TEMA Raw Score - T2	83	27.65	8.81	3	40	68	38.52	7.18	16	59
KeyMath3 Raw Composite Score - T1	89	11.17	5.01	2	21	73	17.48	6.21	4	32
KeyMath3 Raw Composite Score - T2	84	17.00	5.51	2	30	68	27.49	7.74	12	54
KeyMath3 Raw Geometry Score - T1	89	6.04	3.06	0	13	73	7.90	3.01	0	16
KeyMath3 Raw Geometry Score - T2	84	8.23	2.95	0	15	68	11.01	3.03	5	18
KeyMath3 Raw Measurement Score - T1	89	1.28	1.31	0	4	73	3.26	2.15	0	8
KeyMath3 Raw Measurement Score - T2	84	2.88	1.81	0	7	68	6.21	3.32	0	16
KeyMath3 Raw Numeracy Score - T1	89	3.84	1.81	0	8	73	6.11	2.47	2	14
KeyMath3 Raw Numeracy Score - T2	84	5.89	2.04	2	12	68	10.03	2.81	4	21

Table 2. Correlations Among Study Variables

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.
1. First Grade	**1**																		
2. Male	-0.04	**1**																	
3. Offered Treatment	0.07	0.09	**1**																
4. EF T1	0.15	-0.09	-0.14	**1**															
5. Design Copy T1	**0.47**	**-0.18**	-0.02	**0.33**	**1**														
6. WJ App. Probs. T1	**0.56**	**-0.17**	0.00	**0.42**	**0.50**	**1.00**													
7. WJ App. Probs. T2	**0.58**	-0.04	0.07	**0.43**	**0.56**	**0.78**	**1.00**												
8. TEMA T1	**0.65**	-0.14	0.01	**0.40**	**0.53**	**0.77**	**0.71**	**1.00**											
9. TEMA T2	**0.56**	-0.11	0.02	**0.45**	**0.57**	**0.74**	**0.77**	**0.78**	**1.00**										
10. KM3 Composite T1	**0.48**	**-0.16**	0.02	**0.41**	**0.53**	**0.78**	**0.67**	**0.76**	**0.72**	**1.00**									
11. KM3 Composite T2	**0.47**	-0.05	-0.01	**0.46**	**0.61**	**0.78**	**0.83**	**0.81**	**0.80**	**0.80**	**1.00**								
12. KM3 Geometry T1	**0.29**	**-0.18**	0.06	**0.27**	**0.40**	**0.61**	**0.50**	**0.56**	**0.55**	**0.85**	**0.60**	**1.00**							
13. KM3 Geometry T2	**0.50**	-0.16	-0.02	**0.33**	**0.53**	**0.65**	**0.67**	**0.61**	**0.65**	**0.67**	**0.85**	**0.56**	**1.00**						
14. KM3 Measurement T1	**0.65**	-0.08	-0.02	**0.44**	**0.47**	**0.68**	**0.62**	**0.66**	**0.62**	**0.79**	**0.71**	**0.45**	**0.54**	**1.00**					
15. KM3 Measurement T2	**0.61**	-0.09	0.09	**0.47**	**0.51**	**0.70**	**0.73**	**0.73**	**0.69**	**0.70**	**0.89**	**0.49**	**0.61**	**0.67**	**1.00**				
16. KM3 Numeracy T1	**0.42**	-0.04	-0.03	**0.35**	**0.48**	**0.68**	**0.60**	**0.72**	**0.66**	**0.85**	**0.71**	**0.53**	**0.57**	**0.64**	**0.64**	**1.00**			
17. KM3 Numeracy T2	**0.54**	-0.09	-0.11	**0.41**	**0.55**	**0.70**	**0.77**	**0.78**	**0.76**	**0.72**	**0.89**	**0.52**	**0.60**	**0.65**	**0.74**	**0.67**	**1.00**		
18. Composite Math T1	**0.61**	**-0.17**	0.01	**0.45**	**0.57**	**0.92**	**0.79**	**0.92**	**0.81**	**0.92**	**0.86**	**0.73**	**0.70**	**0.77**	**0.77**	**0.81**	**0.80**	**1.00**	
19. Composite Math T2	**0.63**	-0.12	0.01	**0.48**	**0.62**	**0.82**	**0.93**	**0.82**	**0.92**	**0.79**	**0.94**	**0.59**	**0.78**	**0.70**	**0.83**	**0.71**	**0.87**	**0.88**	**1.00**

Notes: Correlations were calculated using listwise deletion.

First Grade coded as 0 = Kindergarten, 1 = First Grade; Male coded as 0 = Female, 1 = Male; Offered Treatment coded as 0 = not offered treatment, 1 = offered treatment; EF = Executive Function Cored Domain Score; T1 = Time 1; WJ App. Probs. = Woodcock Johnson Applied Problems; TEMA = Test of Early Mathematics Ability; KM3 = KeyMath3.

Bold indicates significant correlation with 95% confidence level ($p < .05$).

Table 3. Regression Models for Composite Mathematics – Concurrent and Longitudinal Models

Estimate	Concurrent Models - Composite Mathematics T1						Longitudinal Models - Composite Mathematics T2					
(SE)	Model 1A		Model 1B		Model 1C		Model 2A		Model 2B		Model 2C	
First Grade	1.22	***	0.87	***	0.87	***	0.30	**	0.27	**	0.27	**
	(0.11)		(0.14)		(0.14)		(0.09)		(0.10)		(0.10)	
Male	-0.29		-0.17		-0.17		0.05		0.07		0.08	
	(0.14)		(0.13)		(0.13)		(0.06)		(0.07)		(0.07)	
Received Treatment	-0.04		0.05		0.05		-0.03		0.01		0.01	
	(0.10)		(0.09)		(0.09)		(0.09)		(0.08)		(0.08)	
Composite Mathematics T1							0.86	***	0.74	***	0.74	***
							(0.06)		(0.06)		(0.06)	
EF T1			0.28	***	0.28	***			0.11	**	0.11	**
			(0.06)		(0.06)				(0.04)		(0.04)	
VMI T1			0.26	***	0.26	***			0.15	**	0.14	**
			(0.07)		(0.07)				(0.05)		(0.05)	
EF X VMI T1					0.00						-0.03	
					(0.06)						(0.05)	
R^2	0.39	***	0.55	***	0.55	***	0.79	***	0.82	***	0.82	***
Wald Statistic			28.52	***	0.00				16.82	***	0.54	

NOTES: EF = Executive Function Core Domain Score; VMI = Visuomotor Integration (Design Copy); T1 = Time 1; T2 = Time 2. The Wald statistic shown for each model tests whether parameters existing in the respective model, but not in the model presented immediately to its right, add significant variance explained in the predictor.

Supplemental Tables

Table S4. Concurrent Regression Models – Individual Mathematics Assessments

Estimate	WJ - Applied Problems			TEMA			KeyMath3 - Composite		
(SE)	Model 1D	Model 1E	Model 1F	Model 1G	Model 1H	Model 1I	Model 1J	Model 1K	Model 1L
First Grade	1.11 ***	0.41 ***	0.42 ***	1.27 ***	0.49 ***	0.48 ***	0.96 ***	0.69 ***	0.69 ***
	(0.11)	(0.07)	(0.07)	(0.11)	(0.07)	(0.07)	(0.10)	(0.11)	(0.11)
Male	-0.30 *	-0.10	-0.09	-0.23	-0.06	-0.08	-0.28	-0.16	-0.17
	(0.12)	(0.06)	(0.06)	(0.16)	(0.08)	(0.08)	(0.15)	(0.17)	(0.16)
Received Treatment	-0.04	0.02	0.03	-0.06	0.01	0	0	0.01	0
	(0.11)	(0.06)	(0.05)	(0.11)	(0.05)	(0.06)	(0.10)	(0.13)	(0.13)
EF T1		0.27 ***	0.28 ***		0.24 ***	0.23 ***		0.31 ***	0.31 ***
		(0.06)	(0.06)		(0.06)	(0.06)		(0.08)	(0.07)
VMI T1		0.2 *	0.2 *		0.22 **	0.22 **		0.2 **	0.2 **
		(0.080	(0.09)		(0.07)	(0.07)		(0.07)	(0.07)
EF X VMI T1			-0.08			0.1			0.05
			(0.06)			(0.07)			(0.05)
R^2	0.34 ***	0.46 ***	0.47 ***	0.42 ***	0.53 ***	0.54 ***	0.25 ***	0.41 ***	0.42 ***
Wald Statistic		22.9 ***	1.55		20.46 ***	2.45		29.95 ***	0.07
	KeyMath3 - Geometry			KeyMath3 - Measurement			KeyMath3 - Numeracy		
	Model 1M	Model 1N	Model 1O	Model 1P	Model 1Q	Model 1R	Model 1S	Model 1T	Model 1U
First Grade	0.57 ***	0.26	0.27	0.99 ***	0.69 ***	0.69 ***	0.95 ***	0.62 ***	0.6 ***
	(0.12)	(0.14)	(0.14)	(0.11)	(0.11)	(0.11)	(0.12)	(0.14)	(0.13)
Male	-0.34 *	-0.24	-0.19	-0.27	-0.16	-0.17	-0.06	0.06	0.02
	(0.14)	(0.12)	(0.13)	(0.17)	(0.17)	(0.16)	(0.17)	(0.16)	(0.16)
Received Treatment	0.11	0.17	0.2 *	-0.08	0.01	0	-0.08	-0.01	-0.04
	(0.10)	(0.09)	(0.10)	(0.14)	(0.13)	(0.13)	(0.10)	(0.09)	(0.09)
EF T1		0.16 *	0.17 *		0.31 ***	0.31 ***		0.21 **	0.21 **
		(0.07)	(0.07)		(0.08)	(0.07)		(0.07)	(0.06)
VMI T1		0.27 **	0.26 **		0.2 **	0.2 **		0.27 ***	0.28 ***
		(0.08)	(0.09)		(0.07)	(0.07)		(0.07)	(0.07)
EF X VMI T1			-0.17 *			0.05			0.14
			(0.08)			(0.05)			(0.08)
R^2	0.12 ***	0.22 ***	0.24 ***	0.27 ***	0.41 ***	0.42 ***	0.22	0.35 ***	0.36 ***
Wald Statistic		14.9 ***	5.08 *		20.58 ***	0.75		22.92 ***	2.9

NOTES: EF = Executive Function Core Domain Score; T1 = Time 1. The Wald statistic shown for each model tests whether parameters existing in the respective model, but not in the model presented immediately to its right, add significant variance explained in the predictor.

Table S5. Longitudinal Regression Models – Individual Mathematics Assessments

Estimate	WJ - Applied Problems			TEMA			KeyMath3 - Composite		
(SE)	Model 2D	Model 2E	Model 2F	Model 2G	Model 2H	Model 2I	Model 2J	Model 2K	Model 2L
First Grade	0.41***	0.33*	0.33*	0.24*	0.2*	0.18	0.61***	0.52***	0.52***
	(0.13)	(0.16)	(0.16)	(0.10)	(0.09)	(0.09)	(0.11)	(0.11)	(0.11)
Male	0.09	0.13	0.13	-0.06	-0.02	0.03	-0.01	0.02	0.02
	(0.07)	(0.09)	(0.09)	(0.11)	(0.10)	(0.10)	(0.06)	(0.04)	(0.06)
Received Treatment	0.07	0.11	0.10	-0.02	0.02	0.05	-0.14	-0.05	-0.10
	(0.10)	(0.10)	(0.10)	(0.09)	(0.08)	(0.08)	(0.11)	(0.05)	(0.10)
Pretest	0.68***	0.57***	0.57***	0.70***	0.56***	0.60***	0.65***	0.53***	0.53***
	(0.05)	(0.06)	(0.06)	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)
EF T1		0.12*	0.11*		0.15**	0.15**		0.14**	0.14**
		(0.06)	(0.06)		(0.05)	(0.05)		(0.05)	(0.05)
VMI T1		0.17*	0.17*		0.17**	0.16**		0.16**	0.16**
		(0.08)	(0.08)		(0.05)	(0.06)		(0.05)	(0.05)
EF X VMI T1			0.02			-0.16*			0.04
			(0.05)			(0.06)			(0.06)
R^2	0.65***	0.69***	0.69***	0.62***	0.66***	0.69***	0.71***	0.74***	0.74***
Wald Statistic		190.29***	1.09		36.53***	2.06*		69.77***	0.00
	KeyMath3 - Geometry			KeyMath3 - Measurement			KeyMath3 - Numeracy		
	Model 2M	Model 2N	Model 2O	Model 2P	Model 2Q	Model 2R	Model 2S	Model 2T	Model 2U
First Grade	0.58***	0.35***	0.35***	0.6***	0.53***	0.53***	0.83***	0.74***	0.74***
	(0.09)	(0.09)	(0.10)	(0.17)	(0.16)	(0.16)	(0.15)	(0.15)	(0.15)
Male	0.05	0.12	0.12	-0.04	0.00	0.00	-0.17	-0.11	-0.13
	(0.13)	(0.11)	(0.11)	(0.09	(0.10)	(0.11)	(0.09)	(0.09)	(0.09)
Received Treatment	-0.19	-0.14	-0.14	-0.25*	-0.20	-0.21	0.11	0.14	0.13
	(0.17)	(0.16)	(0.16)	(0.10)	(0.11)	(0.11)	(0.08)	(0.08)	(0.09)
Pretest	0.49***	0.39***	0.39***	0.52***	0.39***	0.38***	0.47***	0.37***	0.35***
	(0.06)	(0.07)	(0.07)		(0.07)	(0.08)	(0.08)	(0.08)	(0.08)
EF T1		0.10	0.10		0.20**	0.20**		0.17**	0.17**
		(0.07)	(0.07)		(0.07)	(0.07)		(0.06)	(0.06)
VMI T1		0.26***	0.26***		0.13*	0.14*		0.14*	0.15*
		(0.06)	(0.06)		(0.07)	(0.07)		(0.07)	(0.07)
EF X VMI T1			0.00			0.01			0.07
			(0.06)			(0.08)			(0.07)
R^2	0.40***	0.47***	0.47***	0.52***	0.57***	0.57***	0.59***	0.64***	0.65***
Wald Statistic		85.95***	3.43		59.02***	0.73		61.56***	0.15

NOTES: EF = Executive Function Core Domain Score; T1 = Time 1. The Wald statistic shown for each model tests whether parameters existing in the respective model, but not in the model presented immediately to its right, add significant variance explained in the predictor.

Table S6. Sensitivity Analyses: Testing Interactions Between Treatment Condition and EF and VMI for Time 2 Mathematics

Estimate (SE)	Composite Mathematics Model 3A	Composite Mathematics Model 3B
Treatment X EF	-0.04	-0.04
	(0.04)	(0.04)
Treatment X VMI	-0.01	-0.00
	(0.03)	(0.04)
Treatment X EF X VMI		0.07
		(0.04)
R^2	0.82	0.82
Wald Statistic	1.14	3.22
Δ df	2	3

	WJ - Applied Problems		TEMA		KeyMath3 Composite	
	Model 3C	Model 3D	Model 3E	Model 3F	Model 3G	Model 3H
Treatment X EF	0.06	0.06	-0.12 *	-0.12 *	-0.03	-0.03
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Treatment X VMI	-0.03	-0.03	0.05	0.05	-0.01	-0.01
	(0.06)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)
Treatment X EF X VMI		-0.02		0.12		0.059
		(0.05)		(0.06)		(0.06)
R^2	0.69	0.69	0.66	0.70	0.74	0.74
Wald Statistic	0.77	1.04	3.79	7.27	0.42	1.46
Δ df	2	3	2	3	2	3

	KeyMath3 Geometry		KeyMath3 Measurement		KeyMath3 Numeracy	
	Model 3I	Model 3J	Model 3K	Model 3L	Model 3M	Model 3N
Treatment X EF	-0.02	-0.02	-0.03	-0.03	-0.01	-0.02
	(0.08)	(0.07)	(0.08)	(0.08)	(0.09)	(0.08)
Treatment X VMI	-0.05	-0.05	0.02	0.02	0.03	0.03
	(0.06)	(0.06)	(0.08)	(0.08)	(0.05)	(0.05)
Treatment X EF X VMI		0.063		0.004		0.082
		(0.07)		(0.09)		(0.09)
R^2	0.47	0.47	0.57	0.57	0.64	0.65
Wald Statistic	1.31	2.83	0.22	0.30	0.28	0.32
Δ df	2	3	2	3	2	3

NOTE: Coefficients for covariates, Time 1 mathematics, Time 1 EF, Time 1 VMI computed for all models but not shown here. Coefficients for Time 1 EF X VMI computed for Models 3B, 3D, 3F, 3H, 3J, 3L, and 3N but not shown here.

Figures

Figure 1. Compensatory Effect between EF and VMI for Time 1 KeyMath3 Geometry.

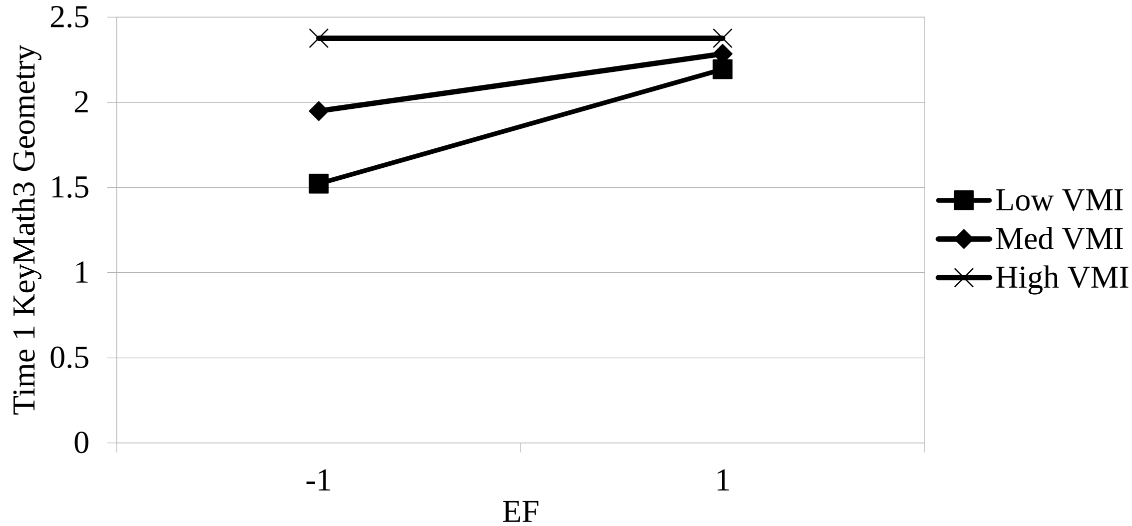
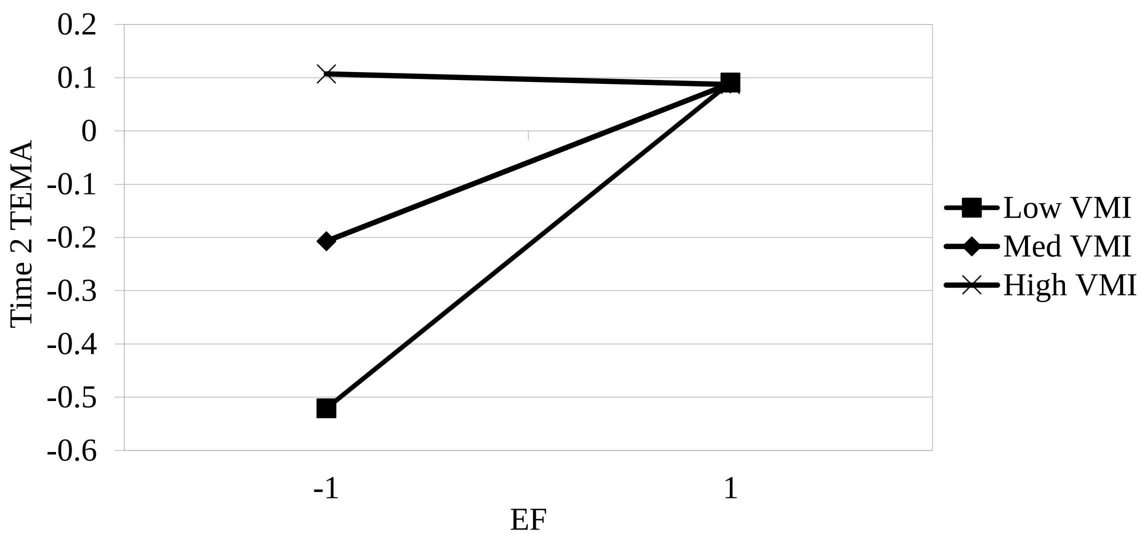


Figure 2. Compensatory Effect between EF and VMI for Time 2 TEMA.



MANUSCRIPT TWO: Family Stress Processes and Children's Self-Regulation

Chelsea A. K. Duran[1], Elizabeth A. Cottone[1], Erik A. Ruzek[1], Andrew J. Mashburn[2], and David W. Grissmer[1]

1. Center for Advanced Study of Teaching and Learning, Curry School of Education, University of Virginia, PO Box 800784, Charlottesville, VA 22908-0784, USA.

2. Department of Psychology, College of Liberal Arts and Sciences, Portland State University, 1721 SW Broadway, Portland,OR 97207-0751

Citation:

Duran, C. A. K., Cottone, E. A., Ruzek, E. A., Mashburn, A. J., Grissmer, D. W. (Under Review). Family Stress Processes and Children's Self-Regulation. *Child Development*

Acknowledgements:

The authors sincerely thank the schools, teachers, and families who participated in this research and without whom this study would not have been possible.

The research reported here was supported by awards from by the Institute of Education Sciences, U.S. Department of Education, through Grant #R305A110703; by the Social Innovation Fund of the Corporation for National and Community Service, managed privately by the Edna McConnell Clark Foundation; and by the Institute of Education Sciences, U.S. Department of Education, through Grant #R305B090002 to the University of Virginia. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Abstract

Economic hardship can affect children's development through child-caregiver interactions, which may mediate cascading effects of other family stress processes. This study examined, simultaneously, the relations of financial strain, caregiver general stress, and child-caregiver conflict – each measured at two time points – with child self-regulatory outcomes in a high-poverty sample (age 5-7 years; *n* = 343). Increase in child-caregiver conflict mediated negative relations between other processes and development of executive function. In contrast, only increase in financial strain had direct, negative association with development of delay of gratification, and did not significantly mediate relations between any other process and children's outcomes. Results have implications for understanding effects of family stress on self-regulatory outcomes, and for interventions with low-income families.

Family Stress Processes and Children's Self-Regulation

Introduction and Literature Review

Family economic hardship can alter the trajectory of children's development, often through the mediating pathway of family stress (Conger et al., 2002; Conger et al, 2010; McLoyd, 1998; Mistry & Wadsworth, 2011). The family stress model (FSM; Conger et al., 2002; Mistry et al., 2002) describes how economic hardship relates to children's outcomes indirectly through family stress processes and child-caregiver interactions. The FSM is well established for social and emotional outcomes, but little is known about its applicability to self-regulatory outcomes, which are sensitive to poverty-related risk and stress exposure (Blair & Raver, 2012; Evans, 2003; Evans & Kim, 2013). Also under studied are contributions of overall levels, versus changes, among family stress processes and to child outcomes (Conger, Conger, & Martin, 2010).

This study examines relations between economic risk and caregiver reports of family stress processes (i.e., caregiver financial strain, caregiver general stress, and child-caregiver conflict) on direct measures of children's self-regulation (executive function (EF) and delay of gratification). Focused on a sample of mostly Black children and caregivers from low-income families, and guided by the FSM, this short-term, prospective study considers unique contributions of family stress processes at two time points on development in child outcomes over the course of one year (from the beginning of kindergarten to the beginning of first grade).

The Family Stress Model (FSM)

Children's interactions within their most proximal contexts, especially their families, are powerful drivers of development, yet factors external to the family can also affect children by altering family processes (Bronfenbrenner, 1986). Families in poverty disproportionately face stressors (e.g., unmet material needs and overcrowding; Evans, 2004), which can impact the development of young children by disrupting family dynamics (McLoyd, 1998). The FSM posits that economic hardship heightens caregivers' financial strain (i.e., perceived sufficiency of current financial resources) and general stress (i.e., perceived degree to which life is stressful, uncontrollable, or unpredictable). These stress processes, in turn, negatively affect children by increasing conflict in family relationships and child-caregiver interactions (Conger et al., 2002; Mistry et al., 2002).

Major aspects of the FSM have been validated across diverse samples, including predominately Black families residing in rural localities and small cities (Conger et al., 2002) and Mexican American families (Parke et al., 2004). Yet, the original model was developed based on experiences of poor, White families living in rural areas (Conger & Elder, 1994), and some aspects of the model may operate differently across racial/ethnic minority groups. For example, material hardship (e.g., food insecurity and residential instability) is related to higher stress in White and Latino parents, and even more so for Black parents. In turn, parents' stress exhibits stronger, negative relations with positive parenting behaviors for Black, than for White, parents (Raver et al., 2007). Further, neighborhood factors – which differ between rural and urban settings – influence many FSM-related processes (see Cassells & Evans, 2017). Therefore, the demographic composition of the present sample – comprising primarily Black, along with some Latino, children and caregivers living in an urban area – is an important consideration.

The FSM has been replicated for children's social and emotional outcomes (see Conger et al., 2010), and even school readiness (Anderson, 2015). But, as discussed further in the next section, relatively few studies have examined the simultaneous effects of these multiple family stress for directly measured, self-regulatory outcomes in young children.

Self-Regulation in Early Childhood and Sensitivity to Hardship and Stress

Self-regulatory outcomes, including EF and delay of gratification, are crucial predictive factors for adaptive academic and social-emotional functioning (Blair & Razza, 2007; McClelland et al., 2014; Moffit et al., 2011; Ursache, Blair, & Raver, 2012). EF refers to cognitive abilities supporting conscious control of thoughts and behavior (i.e., attention, inhibitory control, working memory, and cognitive flexibility; Cameron Ponitz, McClelland, Matthews, & Morrison, 2009). Delay of gratification refers to proclivity to choose “larger-later,” over “smaller-sooner,” rewards, and is generally assumed to reflect self-control (Mischel & Ayduk, 2011). Both EF and delay of gratification develop substantially in early childhood (Best, Miller, & Jones, 2009; Lengua et al., 2015), and improvements in EF in the first year of elementary school are predictive of later academic outcomes (McClelland et al., 2014). At the same time, decision-making heuristics first begin to develop in elementary school (Jacobs & Klaczynski, 2002), which may influence children’s choices to delay gratification (or not). Thus, the early elementary school years may constitute an important and interesting developmental period for these skills.

Recent theory suggests that effects of chronic, poverty-related stress are best viewed as functional adaptations, as opposed to deficiencies, as they support thriving within the environments which shaped them (Wadsworth, 2015). Children’s self-regulatory abilities in particular develop adaptively based on early experiences (Blair & Raver, 2012). In the home, inconsistent routines and unpredictable material provisions may reinforce reflexive (i.e., impulsive or preemptive) behaviors, such as eating treats immediately because they might not be available later. Relative to reflective behaviors (e.g., self-restraint), reflexive behaviors have historically been viewed as deficient and are disadvantageous in some contexts (e.g., school; Blair & Raver, 2012; Finegood & Blair, 2017), but may be advantageous and reflect greater self-regulatory abilities in the face uncertainty or stress (Lee & Carlson, 2015; Blair & Raver, 2012). The adaptive nature of self-regulation development may explain why children from low-income homes – who disproportionately experience unpredictability and unmet needs (Evans, 2004; Evans, Gonnella, Marcynyszyn, Gentile, & Salpekar, 2005) – tend to develop weaker EF and delay gratification less often compared to their higher-SES counterparts (Evans & English, 2002; Evans & Rosenbaum, 2008; Hackman, Gallop, Evans, & Farah, 2015; Noble, Norman, & Farah, 2005). Hence, the importance of examining relations among family stress processes and self-regulatory development in children from low-income families (Finegood & Blair, 2017).

Economic hardship, family stress processes, and EF. Poverty-related risk and stressor exposure (e.g., low maternal education, residential mobility, neighborhood violence), along with physiological stress indicators, partially explain SES-related differences in EF development (Blair et al., 2004; Blair et al., 2011; Blair & Raver, 2012; Evans & Schamberg, 2009; LiGrining, 2007). At the same time, theory and empirical research implicate caregiving as a mediator of the link between economic hardship and development of EF (e.g., Blair & Raver, 2012; Evans & Kim, 2013; Finegood & Blair, 2017). For instance, positive and negative aspects of parenting in infancy mediate relations from maternal education and income-to-needs ratio, respectively, to children's EF at age 3 (Blair et al., 2011). However, whether caregiving also mediates relations from subjective aspects of SES and parental well-being to children's EF is not clear.

Caregivers' perceived financial strain and general stress may also contribute to children's EF development, independent of child-caregiver interactions. Children in poverty are disproportionately and chronically exposed to environmental stressors, such as chaos and overcrowding (Evans, 2004), which could, in turn, affect EF development (LiGrining, 2007; Evans & Kim, 2013). The strong association between years spent in childhood poverty and working memory at age 17 is mediated by physiological stress indicators at ages 9 and 13, whereas maternal stress, sensitivity, and parenting styles do not always mediate this relation (Evans & Schamberg, 2009). Thus, poverty-related stressors could affect children's EF development, independent of child-caregiver interactions. Along these lines, and in the same sample as in Blair et al. (2011), Raver, Blair, and Willoughby (2013) found that *both* chronic exposure to objective economic risk factors *and* financial strain were uniquely related to children's EF at age 4.

Notably, the studies just discussed included only one or two FSM-relevant family processes, leaving others as potential confounds in the observed relations. For instance, it is unclear whether parenting factors (as in Blair et al., 2011) and financial strain (as in Raver et al., 2013) would be uniquely related to children's EF if included in the same model. Clarifying unique relations between family stress processes and children's EF could inform interventions targeting low-income caregivers' stress and children's behavior (e.g., Neville et al., 2013). The FSM posits that caregiving mediates effects of other family stress processes on children's outcomes and is therefore a key leverage point for intervention, as opposed to more distal family stress processes. The present study is the first to explicitly test this feature of the FSM for children's EF.

Economic hardship, family processes, and delay of gratification. As with EF, common explanations for the links between SES and delay of gratification performance highlight risk and stressor exposure (Blair & Raver, 2012; Evans & Kim, 2013). However, in contrast with EF, research has drawn inconsistent conclusions regarding the relations of risk exposure and physiological stress indicators with delay of gratification (Evans, 2003; Lengua, Zalewski, Fisher, & Moran, 2013). Thus, the pathways leading to income-related differences in delay of gratification may not be the same as those for EF.

Several studies have found associations between “co-regulating” parenting behaviors or mother-child connectedness and delay of gratification (Houck & Lecuyer-Maus, 2004; Lengua, Honorado, & Bush, 2007; LiGrining, 2007). Should these reflect causal relations, then more distal processes to the child (e.g., financial strain and general stress) could indirectly affect children’s delay of gratification by disrupting child-caregiver interactions. However, caregiver financial strain and/or caregiver stress have not typically been considered in such studies and, therefore, whether child-caregiver interactions mediate their effects on delay of gratification development is unknown.

As with EF, aspects of economic hardship and/or caregiver financial strain may play a direct role in children’s choices to delay gratification. Delaying gratification depends upon context-dependent decision-making processes, and may not be adaptive when long-term rewards are unlikely or uncertain (Fawcett, McNamara, & Houston, 2012; Rosati & Stevens, 2009). Indeed, children’s delay of gratification is negatively affected when children’s perceived likelihood of receiving “later rewards” is experimentally decreased (Kidd et al., 2013; Lee & Carlson, 2015). Because children from low-income families disproportionately face unpredictability and unmet needs, they may be less motivated to decline an immediate, tangible reward than their more advantaged peers (Kidd et al., 2013). Compared to objective economic risks, caregivers' subjective perceptions of financial strain are more closely related to family functioning and children's outcomes (Mistry, Biesanz, Taylor, Burchinal, & Cox, 2004), though the unique influence of each of these factors on children's delay of gratification has not been studied. Thus, in this study we test whether economic hardship and/or caregiver financial strain are related to children's delay of gratification, independent of child-caregiver conflict.

Black children from low-income families may be particularly sensitive to uncertainty of promised rewards when choosing between immediate and delayed gratification. Black, but not White, elementary-aged children from low-income families delay gratification less often when tested by a White, rather than Black, experimenter (Strickland, 1972). Even when children and experimenters are of the same race, Black children express less trust in experimenters than White children (Price-Williams & Ramirez, 1974). Thus, investigating relations of economic hardship and family stress processes with children's delay of gratification may be particularly interesting in the present sample, comprising predominantly Black children from low-income families.

The Importance of Considering Change in Family Stress Processes Over Time

Chronic exposure to poverty profoundly shapes children's self-regulatory development (Blair & Raver, 2012), and is associated with more adverse effects than are fluctuations in and out of poverty (Evans, 2004; McLoyd, 1998). For instance, duration of exposure to economic risk and caregiver financial strain in children's first four years of life each exhibit unique relations with EF at age 4, even after accounting for the severity of economic risks faced by the family during infancy (Raver et al., 2013). Yet, quasiexperimental evidence suggests short-term changes in income affect the quality of family relationships and some child outcomes (Costello, Compton, Keeler, & Angold, 2003). Similarly, children who have experienced *either* persistent *or* transitory poverty score lower on cognitive tests than children who have never experienced poverty (Duncan et al., 1984; McLoyd, 1998). Thus, both overall levels and changes in poverty-exposure and income seem to influence children's social, emotional, and cognitive development.

However, research establishing the FSM – which provides a theory for how poverty-exposure affects children and can therefore inform interventions – has been largely based on cross-sectional data (Conger et al., 2010). A recent study utilizing longitudinal data validated the FSM, but still only measured each construct at a single, separate time point (Neppl, Senia, & Donnellan, 2016). Similarly, the bulk of the EF-focused literature discussed earlier (e.g., Blair et al., 2011; Raver et al., 2013) comprises studies measuring EF at a single time point, severely limiting the warrant for causal interpretation of observed associations. Static and changing aspects of parenting relate significantly to EF development over time (Blair, Raver, & Berry, 2014), but whether these are robust to the inclusion of other family stress processes is unclear. Thus, whether key features of the FSM are robust to consideration of both baseline levels and changes in multiple processes simultaneously is a focus of the present study.

Generally speaking, strong relations among changes in family stress processes and children's outcomes would be consistent with the notion of a causal cascade, and would be informative from an intervention perspective. On the other hand, weak or insignificant relations among changes in family stress processes and children's outcomes may indicate that some processes operate as stable or immutable family traits, and/or are relatively unresponsive to changes in other processes.

Present Study

In a sample of mostly Black children and caregivers from low-income families, we explore the direct and indirect relations from two indicators of economic risk (i.e., mother's education and family receipt of public assistance other than free/reduced price school lunch) and family stress processes (i.e., caregiver financial strain, general stress, and child-caregiver conflict) to children's performance and improvement in EF and delay of gratification from the beginning of kindergarten (age 5-6) to the beginning of first grade (age 6-7). Examining each family stress process at both time points disentangles the statistical contributions of baseline levels and changes in each process to improvement in children's outcomes.

In general, and in line with the FSM, we expected greater economic risk to relate to higher family stress levels, which would at least partially mediate the relations from risk to weaker self-regulation and/or less growth in self-regulatory outcomes. We expected significant relations between financial strain and both EF and delay of gratification, and only partial mediation of these through child-caregiver conflict. We also expected direct relations from either financial strain or general stress–potential proxies for other unmeasured stressors, such as overcrowding–to EF, and direct effects of economic risk and financial strain–which could affect children's perceptions of reliability of rewards–on delay of gratification.

Given research focused on long-term cumulative risk exposure and canalization of self-regulation development over time (Blair & Raver, 2012; Evans & Kim, 2013), we expected that contributions of baseline levels of family processes on child outcomes would be strong and in the negative direction. We also expected that increases in family stress processes would relate to less improvement in children's outcomes, but perhaps not as consistently and/or strongly as contributions of baseline levels.

While the focus of our study is on how family stress processes relate to children's self-regulatory development, our analyses also include, and therefore test, relations between family stress processes between two points in time. Past FSM research has been based on data including measurement of each stress process at a single time point (usually, concurrently; Conger et al., 2010). This disallows clear predictions regarding how, for example, level of caregivers' financial strain at one time point might affect changes in their general stress over the course of the following year, above and beyond its effect on concurrent general stress and the effect of later financial strain. Thus, we make no firm hypotheses regarding how each process might predict change in the others.

Method

Participants

Participants in this study (*n* = 343) were a subset of 371 five- to seven-year-old children and their caregivers participating in a randomized control trial evaluation of an after-school, social-emotional learning program and attending one of four elementary schools in an urban school district in the Southeastern United States. Three cohorts of children were recruited over the course of three consecutive school years, with each cohort's participation beginning in the summer prior to or the beginning of kindergarten (average age 5.4 years). A majority of caregivers were mothers (92%); others were fathers (3%), grandmothers (3%), and other family members (e.g., aunts). Twenty-six children were excluded from analysis because their caregiver-respondents differed between time points, which would complicate interpretation of changes in caregiver reports. Two children who were missing all data except demographics were also excluded. At Time 1, 96% of children were receiving free- or reduced-price lunch and 89% of caregivers identified as Black (6% as Latino; see Table 1), which roughly mirrors the broader population served by the schools in their district.

Procedures and Measures

All constructs considered in the present study were measured at two time points (Time 1: summer prior to or beginning of kindergarten; Time 2: summer prior to or beginning of first grade). Economic risk indicators from Time 1 only were used in our analyses, as these indicators were largely stable between time points.

Children's self-regulatory outcomes. During each summer of the study, children were offered the option to attend a free, week-long "summer camp" at their schools, during which they were administered three, 15- to 20-minute batteries of direct assessments by trained research assistants. The measures of self-regulation used herein were assessed during two separate batteries. Children who did not attend summer camp were assessed during the first half of the school year in a quiet space at their school.

EF was measured using Head-Toes-Knees-Shoulders (HTKS; Cameron Ponitz et al., 2009), which requires children to perform sets of paired rules for doing the "opposite" of verbal directions (e.g., "When I say touch your toes, touch your head"). Part one of three includes only one such rule; part two adds a second rule (e.g., "When I say touch your shoulders, touch your knees"); and in part three, rules switch. HTKS simultaneously taps multiple aspects of EF, in that it requires children to pay attention to directions, inhibit the natural response to follow directions directly, maintain multiple rules in mind, and flexibly switch response rules between parts. Each of the three parts of the assessment has 10 items and is preceded by six practice items; feedback regarding performance is given on practice items only. The HTKS composite score is a sum of 96 possible points across 48 items (2, 1, and 0 points for correct, self-corrected, and incorrect responses, respectively). Internal reliability (Cronbach's alpha) in the current sample was at least α = .89 at each time point.

Delay of gratification was measured using a choice delay task, which presented children with a series of six choices between small rewards (e.g., 1 sticker), received immediately, or larger rewards (e.g., 6 stickers), which they would have access to later in the day (Hongwanishkul, Happaney, Lee, & Zelazo, 2005). Three items used stickers as rewards, three items used candy; for each reward-type, ratios between smaller and larger rewards were 1:2, 1:4, and 1:6. Internal reliability in the current sample was at least α = .83 at each time point.

Economic risk and family stress processes. Caregiver questionnaires including inquiries about demographic information and family stress processes were administered in-person during summer camp. When caregivers were not available during the camp, questionnaires were administered over the phone. Composite scores were calculated (average scores across all items) for each family stress process scale at each time point.

Economic Risk. Our analyses include two indicators for economic risk: mother's education and whether the family was receiving public assistance other than free/reduced price lunch (e.g., food stamps, public housing). Mother's education was originally reported on an 8-category scale, but in our analyses, this variable was dichotomized to indicate whether the mother had dropped out of high school (29% of sample).

Family Stress Processes. Caregiver's financial strain was measured with a three-item, 5-point, Likert-style scale focused on difficulty of living on current income and ability to maintain standard of living (e.g., “How difficult is it for you to live on your total household income right now?” and “In the next two months, how much do you anticipate having to reduce your standard of living to the bare necessities of life?"; Vinokur, Price, & Caplan, 1996). Internal reliability in the current sample was α = .75 at both time points.

Caregiver general stress was measured using the 13-item, 5-point perceived stress scale focused on ability to cope with life challenges, as well as levels of anxiety and stress (Cohen, Kamrack, & Mermelstein, 1983). Example items include: “In the last month, how often have you found that you could not cope with all the things that you had to do?" and “In the last month, how often have you felt nervous and ‘stressed’?" Internal reliability in the current sample was at least α = .82 at both time points.

Child-caregiver conflict was measured with the 8-item, 5-point Likert-style conflict subscale from the Child Parent Relationship Scale, which assesses child and caregiver anger and relational, or parenting, stress. Example items include “My child and I always seem to be struggling with each other" and “Dealing with my child drains my energy" (Pianta, 1992). Internal reliability in the current sample was α = .74 at both time points.

Analytic Plan

For each outcome, the study goals were achieved simultaneously through a structural equation modeling framework using a single path model carried out in Mplus v. 7.0 (Muthén & Muthén, 2005) and using Full Information Maximum Likelihood (FIML) estimation to address missing data (see more details on missing data below).

Figure 1 depicts the path models estimated for each outcome. Family economic risk indicators were free to predict all family processes and child outcomes at both time points.

Direction of relations among family processes and child outcomes at each time point were constrained according to the FSM, with financial strain contributing to general stress and child-caregiver conflict and general stress contributing to child-caregiver conflict. Within each time point, all three family processes were free to predict to child outcomes. Finally, all Time 1 variables were free to predict all Time 2 variables. The fit of each model was assessed using Tucker-Lewis index (TLI) and comparative fit index (CFI) greater than 0.95 and root-mean-square (RMSEA) less than or equal to 0.06 (Hu & Bentler, 1998).

To address our first research goal, we examined the significance of estimates for path loading parameters directly from family economic risk, financial strain, general stress, and child-caregiver conflict to child outcome. We also examined the significance of indirect effects of economic risk, financial strain and general stress through other processes for each child outcome. Indirect effects were considered significant if zero fell outside 95% confidence intervals, which were calculated using bootstrapping to account for asymmetry in the distribution of indirect effects (MacKinnon, Lockwood, & Williams, 2004). Our second goal was achieved by evaluating whether family processes from each time point (the first representing initial levels and the second representing change in each process) exhibited unique, significant contributions to improvement in children's outcomes.

Covariates. Children's age at Time 1 assessment and gender were used as covariates, as were study cohort (Cohort 1, 2, and 3) and randomized treatment condition. Those randomized into the control group were coded as "0 = not offered treatment"; those randomized into the treatment group were coded as "1 = offered treatment." Eighteen children were not randomized, but were offered the opportunity to attend the program, and coded as "1 = offered treatment," due to an older sibling already attending.

Missing Data. Each analytic variable was missing between 0 and 31% of cases (see Table 1). Among cases in the analytic sample (*n* = 343), 185 (54%) had complete data across all analytic variables. An additional 75 children (22%) were missing only one type of data (caregiver reports or child direct assessments) at one time point. Our analyses employed FIML estimation, which uses all possible information, produces less biased estimates than deletion methods, and is valid under conditions where data can be assumed to be missing at random after accounting for missing data mechanisms (Acock, 2005). Study attrition accounts for the vast majority of missing data in our sample, evidenced by higher missing data rates at Time 2 than at Time 1, which could introduce selectivity bias. However, missingness was not related to analytic variables, with two exceptions: the third cohort had the highest rates of missing data across all data types and time points, and fewer caregivers of control children responded at Time 1 relative to those of treatment children.

We explored other, non-study variables as candidates for auxiliary variables. Later child direct assessment timing was related to nonresponse by caregivers and, thus, timing of child direct assessments were included as auxiliary variables. Further, so that both models had identical information available for FIML, children's HTKS scores were used as auxiliary variables in the Choice Delay model, and vice versa.

Results

Table 1 provides descriptive statistics for analytic and demographic variables. Financial strain, general stress, and child-caregiver conflict had similar means between Time 1 and Time 2, suggesting no systematic change in the sample over time. In other words, just as many caregiver reports indicated a decrease as those that indicated an increase in each process from Time 1 to Time 2. HTKS scores exhibited floor effects, with a modal score of zero, at Time 1. Average HTKS scores were 28.6 points (SD = 27.20) and 53.75 points (SD = 26.68) at Time 1 and Time 2, respectively. Average Choice Delay scores were 3.16 (SD = 2.20) and 3.56 (SD = 2.22) at Time 1 and Time 2, respectively.

Bivariate correlations among analytic variables, also in Table 1, were generally in the expected direction, when significant. Each caregiver report measure had moderate to high correlation with itself between Time 1 and Time 2 (ranging from .34 to .56), suggesting stability in family stress processes. Further, correlations between family stress processes were positive, within and between both time points. Unexpectedly, HTKS and Choice Delay performance were negatively related at Time 1 (r = .20, p < .05).

Relations Among Economic Risk Indicators and Family Stress Processes

Not surprisingly, given that models were nearly saturated, fit for both the HTKS and Choice Delay models were excellent, with TLI of 1.03 and 1.01, CFI of 1.00 and 1.00, and RMSEA of .00 and .00, respectively.

Relations among economic risk and family stress processes were largely similar between the HTKS and Choice Delay models. For simplicity, we show and report only loadings from the HTKS model here (Figure 2; see supplemental materials for parameters for the Choice Delay model). Mothers having dropped out of high school was positively related to Time 1 general stress in (β = .11, p < .05), and receiving public assistance was positively related to Time 2 financial strain (β = .12, p < .05). Time 1 financial strain was significantly and positively related to Time 1 general stress (β = .51, p < .001), which in turn was significantly and positively related to Time 1 child-caregiver conflict (β = .34, p < .001). At Time 1, financial strain had a significant, specific indirect effect on child-caregiver conflict through general stress (indirect β = .17, CI[.10, .24]; direct-mediated effect β = -.07, CI[-.20, .06], total effect β = .10, CI[-.004,.21]). Time 2 financial strain was significantly and positively related to Time 2 general stress (β = .32, p < .001), which in turn was significantly and positively related to child-caregiver conflict (β = .34, p < .001). Time 2 financial strain had a significant, specific indirect effect on Time 2 child-caregiver conflict through Time 2 general stress (β = .11, CI[.04, .18]; direct-mediated effect β = -.03, CI[-.15, .09] , total effect β = .08, CI[-.04, .20]).

There were also several significant contributions from Time 1 family stress processes to Time 2 family stress processes. First, each family stress process at Time 1 was significantly and positively related to itself at Time 2 (financial strain: β = .25, p < .01; general stress: β = .37, p < .001; and child-caregiver conflict: β = .48, p < .001). Further, Time 1 general stress positively predicted Time 2 financial strain (β = .18, p < .05), and Time 1 child-caregiver conflict positively predicted Time 2 general stress (β = .20, p < .001). A detailed account of indirect relations among family stress processes between time points is not included here, given the present study's focus on relations with children's outcomes. Relevant, indirect relations among family processes are implicated in later sections describing indirect relations between family processes and children's outcomes.

Economic Risk, Family Stress Processes, and Children's EF.

Figure 2 and Table 2 show parameter estimates for direct effects on children's HTKS performance; Table 3 provides estimates and confidence intervals for total, total indirect, and specific indirect effects on HTKS performance. The model explained 9% and 34 % of variance in HTKS performance at Time 1 and Time 2, respectively. Children whose mothers had not graduated from high school performed worse on HTKS at Time 1 than children of high school graduates (β = -.13, p < .05). None of the family stress processes included in the model was related, directly or indirectly, to HTKS at Time 1.

HTKS performance exhibited stability, such that Time 1 performance significantly and positively predicted Time 2 performance (β = .49, p < .001). Neither economic risk indicator was directly related to Time 2 HTKS performance, but mother's high school dropout status had a significant, negative, total indirect effect (β = -.11, CI[-.179, -.041]), with one significant, specific indirect effect through Time 1 HTKS performance (specific indirect = -.06, [CI = -.122, -.007], total effect β = -.08, CI[-.214, .041]).

Among family stress processes, only Time 2 child-caregiver conflict exhibited significant, direct effects on children's Time 2 HTKS performance, with greater conflict associated with worse performance (β = -.23, p < .01). Time 1 child-caregiver conflict had a significant, negative, total indirect effect on Time 2 HTKS performance (total indirect effect β = -.13, CI[-.237,-.025]; total effect β = -.11, CI[-.244, .022]), with significant, specific indirect effects through Time 2 child-caregiver conflict (specific indirect effect: β = -.11, CI[-.197, -.029]) and through Time 2 general stress to Time 2 child-caregiver conflict (specific indirect effect β = -.02, CI[-.031, -.001]).

Increased financial strain and general stress at Time 1 and Time 2 each indirectly related to less improvement in HTKS performance, with child-caregiver conflict at Time 2 always serving as the final mediator. The total indirect effect of Time 1 financial strain on Time 2 HTKS was significant and negative (total indirect effect β = -.11, CI[-.211, -.014], total effect β = .00, CI[-.127, .127]), with one significant, specific indirect effect through Time 1 general stress to Time 1 child-caregiver conflict and finally through Time 2 child-caregiver conflict (β = -.02, CI[-.034, -.003]). Time 2 financial strain did not have a significant total indirect effect on Time 2 HTKS, but had one borderline-significant, negative, specific indirect effect through Time 2 general stress and Time 2 child-caregiver conflict (β = -.03, CI[-.051, .000]). Time 1 general stress did not have a significant total indirect effect on Time 2 HTKS, but had two significant, negative, specific indirect effects: one through child-caregiver conflict at Time 1 and child-caregiver conflict at Time 2 (β = -.04, CI[-.069, -.007]) and another through general stress at Time 2 and child-caregiver conflict at Time 2 (β = -.03, CI[-.057, -.001]). Time 2 general stress had a significant, negative total indirect effect on Time 2 HTKS performance through Time 2 child-caregiver conflict (indirect effect: β = -.08 CI[-.145, -.015]; total effect β = -.09, CI[-.256, .087]).

Economic Risk, Family Stress Processes, and Children's Delay of Gratification.

A figure providing the Choice Delay model, along with a table with all parameter estimates are provided in supplemental materials. Other than covariates, none of the variables in the model was significantly related to children's Time 1 Choice Delay performance, of which only 3% of variance was explained by the model.

The model explained 17% of variance in Choice Delay performance at Time 2, which exhibited some stability, with Time 1 positively predicting Time 2 performance (β = .28, p < .01). Neither economic risk indicator was directly related to Time 2 Choice Delay performance. Among family processes, higher financial strain at Time 2 was directly related to worse Time 2 performance (β = -.16, p < .05). There were no significant, indirect relations between any family process and Choice Delay performance.

Sensitivity analyses.

Sensitivity analyses were run order to test whether participation in the after school program evaluation, from which these data are derive, interfered with or drove our results. For each outcome, a multi-group model was run allowing parameters to differ between treatment groups, followed by a multi-group model in which all path loadings were constrained to be equal between groups. Chi-squared difference tests demonstrated that each constrained model did not have worse fit than its corresponding unconstrained model: $\Delta\chi^2(76) = 71.69$, $p = .62$ for HTKS; $\Delta\chi^2(76) = 70.40$, $p = .53$ for Choice Delay (see supplemental materials for full fit information on each model). Thus, we find no evidence that model parameters differed systematically between treatment and control groups or, in other words, that after-school program participation interfered with or drove our results.

Discussion

Guided by the FSM, the goal of this study was to extend current understanding of relations between economic risk, family stress processes, and development of children's self-regulation in a low-income sample of mostly Black, early elementary-aged children and their caregivers. Past research had studied relations between specific aspects of family risk and stress with children's EF and delay of gratification, but not simultaneously in a single model as we have done. Results suggest differing pathway(s) by which family stress processes relate to the development of each outcome. The present study is also novel in its inclusion of two time data points for each family stress process. Only changes – and not baseline levels – in family processes were directly related to improvement in each outcome. This is consistent with the notion of stress processes exhibiting dynamic influence on children, as opposed to representing unchanging family characteristics.

Relations among Economic Risk Indicators and Family Stress Processes

Our economic risk indicators were not related to baseline financial strain, but financial strain increased among caregivers receiving public assistance, such as public housing vouchers or food stamps. This indicator did not measure *eligibility* for receiving public assistance, and so is not a perfect proxy of relative income disadvantage in this low-income sample. It may be that some of the approximately 20% of families who were not *receiving* public assistance at baseline could still have been *eligible* to receive public assistance. Such families could ostensibly be more strained financially than those receiving assistance; this would in turn weaken any relation between receiving public assistance and concurrent financial strain. Thus, in a sample as homogeneously disadvantaged as ours, it may be necessary to measure economic hardship directly and more precisely (e.g., with income-to-need ratio) to detect effects on concurrent family stress processes.

Mother's high school dropout status was directly and positively related to caregivers' baseline general stress. This is consistent with other research uncovering links between lower educational attainment and greater parenting-related stress in Black mothers (Cardoso, Padilla, & Sampson, 2010), who constituted the vast majority of the present sample. Higher education may secure familial resources – including psychological or social resources – which support coping with hardship (Ross & Van Willigen, 1997).

Direction of modeled relations among family stress processes within time points was taken from the FSM, as directionality cannot be tested from these data; these were positive and significant, as expected, for both baseline levels of and changes in stress processes. Prior research examining multiple FSM processes has typically measured each construct at only one time point, leaving open the possibility that unmeasured family characteristics are confounded in relations. Modeling relations among changes in multiple stress processes over time, as we have done, does not entirely eliminate the possibility of confounds. However, this design limits potential confounds to those which are themselves dynamic or exhibit dynamic and independent relations with each stress process over time.

Baseline general stress and child-caregiver conflict each predicted increases its respective "upstream" process (i.e., financial strain and general stress). Cross-lagged paths among processes over time are unexplored elsewhere in the literature and are not explicitly included in the FSM. The FSM does not necessarily posit that baseline financial strain, for example, should forecast increases in, above and beyond its concurrent influence on, its "downstream" process (i.e., general stress). Instead, the cross-lagged relations from "downstream" processes to changes in "upstream" processes, as we observe, lend support to the notion of bidirectional feedback loops among family processes. For example, general stress of parents may increase financial strain down the road by reducing caregivers' capacity to cope effectively and make sound decisions in the face of financial challenges (as speculated by Conger et al., 2010). Similarly, high baseline child-caregiver conflict could increase caregivers' general stress as a side effect of detrimental effects of conflict on children's social-emotional development and other family dynamics.

Economic Risk, Family Stress Processes, and Children's EF

Mother's education was directly related to children's baseline EF, and indirectly to EF at Time 2 through baseline EF. This is consistent with Raver and colleagues' (2013) observation of direct relations between mother's education and children's EF at age 4 after controlling for financial strain, but inconsistent with Blair and colleagues' (2011) observation that parenting factors mediate relations from mother's education to children's EF at age 3. Our measure captured caregivers' perceptions of conflict in the relationship with their child, whereas Blair and colleagues (2011) directly observed child-parent interactions. Despite differences in measurement, both sets of results are consistent with assertions that education improves families' ability to invest in children's development (Yeung, Linver, & Brooks-Gunn, 2002).

Family processes at baseline did not significantly contribute to children's EF performance at baseline. This was unexpected, given that much of the cross-sectional FSM literature (Conger et al., 2010) has observed effects on children's concurrent social and emotional outcomes. This may have been due to an unexpected floor effect on the HTKS measure at baseline, apparently absent in other studies focused on younger populations (e.g., McClelland et al., 2014), and potentially reflecting the challenges that children in this population face in their EF development.

Children whose caregivers reported an increase in child-caregiver conflict improved less in their EF than children whose caregivers reported a decrease in conflict. Baseline levels of conflict were indirectly related to children's EF development, but only through a stable relationship with Time 2 conflict. This finding replicates prior FSM research positing child-caregiver interactions as a direct influence on children's development. This also extends prior understanding by suggesting a stronger influence of changes in, compared to baseline levels of, child-caregiver interactions on children's EF development.

We found specific, negative, indirect effects of caregiver financial strain and general stress on improvement in children's EF through increases in child-caregiver conflict. These findings are consistent with Raver and colleagues' (2013) finding that caregiver financial strain exhibits longitudinal relations with young children's EF after accounting for economic risk. However, we hypothesized that child-caregiver conflict would only *partially* mediate relations from financial strain and general stress to children's

EF. This hypothesis was based on the notion that the latter two may proxy poverty-related stressors directly affecting EF development. If such stressors were influential, they did not influence caregivers' reports of financial strain or general stress without also being confounded in caregivers' reports of child-caregiver conflict. Thus, our hypothesis may be more viable in a study utilizing a direct measure of parenting behaviors (e.g., as in Blair et al., 2011) as opposed to a caregiver report. Nevertheless, a mediating role for caregiving factors between family stress and children's EF is consistent with the FSM and the notion that caregivers play a key role in shaping self-regulatory modes (Blair & Raver, 2012).

Importantly, relations from initial levels in family stress processes were only indirectly related to improvement in EF through their relations with later family stress processes. This is consistent with evidence that both long-standing status and changes in poverty-related factors affecting family stress processes influence children's outcomes, including EF (Duncan & Magnuson, 2005; Evans, 2004; McLoyd, 1998; Raver et al., 2013), but suggests that contemporaneously measured, rather than previously measured, family stress is more closely related to children's EF. This observation is good news from an intervention perspective: were this not the case, then this might indicate immutability in one or more stress process. Alternatively, and/or additionally, this could reflect the presence of unmeasured, confounding family characteristics generating relations among processes at baseline and with children's EF development. Of course, the present study is observational by design and causal interpretations of these relations are unwarranted, but inclusion of two time points of both predictors and outcomes greatly reduces the likelihood for static family characteristics explaining results.

Economic Risk, Family Stress Processes, and Children's Delay of Gratification

Neither economic risk indicator was related to delay of gratification at either time point. Associations between SES and delay of gratification are strong when utilizing "sustained delay" tasks requiring children to uphold a choice to delay gratification over time (e.g., Evans & English, 2002; Evans & Rosenbaum, 2008), as in the classic "marshmallow test" (Mischel & Ayduk, 2011). In contrast, differences were not found in *choice* delay performance, despite profound differences in EF, between low- and middle-class kindergarteners (Noble, Norman, & Farah, 2005). Thus, our results reinforce the notion that performance on choice delay tasks does not vary by objectively measured factors of class or SES.

None of our family stress processes contributed to children's delay of gratification at baseline, whereas past studies have observed cross-sectional associations between caregiving interactions and children's ability to delay gratification (e.g., Houck & Lecuyer-Maus, 2004; Lengua et al., 2007; LiGrining, 2007). The lack of such a finding in this study could be due to our measures of both child outcomes and caregiving. First, prior research establishing these links primarily used task paradigms other than choice delay (e.g., "snack delay" or "gift delay"; Lengua et al., 2007). Second, our measure of child-caregiver interactions was self-reported, which differs conceptually from observed behaviors used in other studies. Alternatively, we speculate later that other factors – especially or perhaps uniquely influential for Black children from low-income families – could have influenced these expected relations.

Unexpected according to the FSM, but consistent with our hypotheses influenced by a functional adaptation approach (Blair & Raver, 2012; Wadsworth, 2015), we found negative, direct relations between increases in financial strain and improvement in children's delay of gratification. We speculate that the influence of an adaptive decision-making process may account for this relation. Experimental paradigms lowering children's expectations about the certainty of later rewards greatly reduce their tendency to delay gratification (Kidd et al., 2013), especially among children with high cognitive flexibility (a component of EF; Lee & Carlson, 2015). Caregivers experiencing an increase financial strain from one year to the next may, as a result, manage household material resources more strictly or with less predictability. Children may in turn be primed to suspect that "larger-later" rewards may not be certain, and may, in turn, prefer immediate over delayed gratification. This may be especially relevant for the choice delay task paradigm utilized here, which required forfeiting access to "smaller-sooner" rewards during the several hours in which they had to wait for "larger-later" rewards. This design carries an inherent risk of "no reward" for children who perceive that as a possibility.

The "trust" or "expectation" explanation for the relation between financial strain and delay of gratification is speculative, but could have muddled the latter construct – especially during children's first encounter with experimenters. As previously mentioned, trust in experimenters and sensitivity to social context may be uniquely influential for Black children from low-income families, as compared to children from other races from low income families (Price-Williams & Ramirez, 1974; Strickland, 1972). Multi-dimensionality of children's DoG performance at baseline – including trust/expectancy along with self-regulatory ability – may be one tentative explanation for why almost no study variables were related to children's delay of gratification at baseline. This may also explain the unexpected, negative correlation between EF and choice delay at baseline, appearing in all three study cohorts but not at Time 2 in any cohort. Children with strong cognitive flexibility are more likely to adjust their delay of gratification preferences in response to lowered certainty of later rewards (Lee & Carlson, 2015). In the present study, the first encounter may have represented an uncertain scenario for children, perhaps explaining why children with stronger EF to delayed gratification less than those with weaker EF. Children's prior first- and/or second-hand experiences may have diminished concerns regarding certainty of later rewards, along with this unexpected, negative relation, at Time 2. Given the racial composition and extreme economic disadvantage in this sample, it may be the case that for most of these children, choosing not to delay gratification represented an adaptive response at baseline.

Limitations and Directions for Future Research

Because this study utilized secondary data from a larger study not centrally focused on family stress, the measures used here are not identical to prior FSM research, limiting comparisons of our findings with those of prior studies. A formal test of the FSM would have involved direct measures of family economic hardship (e.g., income or income-to-needs ratio), whereas this study utilized two coarse indicators for economic risk. Additionally, we used caregivers' reports of perceptions of conflict with their children, as opposed to directly observed child-caregiver interactions as in prior FSM studies. However, our findings with respect to improvement in children's EF align with what would be expected based on prior FSM research, suggesting this self-reported measure somewhat adequately represented the child-caregiver interaction construct in this study.

Measurement of child outcomes may also be of some concern. First, EF – a multidimensional construct – is most robustly measured with multiple tasks (Willoughby, Wirth, & Blair, 2012), whereas this study used only one task. Further, this single measure exhibited floor effects at Time 1; this was unexpected, and may explain the lack of observed associations with family stress processes at Time 1. Second, EF and delay of gratification were negatively correlated at Time 1, which was unexpected, despite our earlier speculations regarding trust and certainty as an influential dimension in delay of gratification, especially at Time 1. A meta-analysis by Duckworth and Kern (2011) reports a zero-correlation between performance on choice delay tasks and EF tasks, as we observe at Time 2. Both of these measurement concerns might have reduced power to detect effects of family processes on children's self-regulatory outcomes at Time 1; we do not see cause for concern regarding measurement of family stress processes at either time point or of children's outcomes at Time 2.

The path modeling approach utilized here does not warrant uncaveated causal interpretations of the relations uncovered. For instance, our analyses cannot affirm directionality among constructs, in part due to concurrent data collection among constructs. For example, they do not rule out the possibility that difficulties in children's EF development lead to greater child-caregiver conflict, as opposed to the latter affecting the former. However, our interpretations are guided theoretically by the FSM, and the inclusion of each family stress process at more than one time point represents an improvement over prior research in this area (Conger et al., 2010). Thus, these results should only be viewed as a "stepping stone" toward understanding causal effects of family stress processes on children's self-regulatory development.

Interestingly, cross-lagged relations among family stress processes are consistent with the substantive notion of bidirectionality in their influences on each other. On the other hand, relations from "downstream" to changes in "upstream" processes could be statistical artifacts of the saturated nature of the models run and unequal numbers of variables predicting each stress process at Time 2. For instance, baseline child-caregiver conflict predicting an increase general stress, but not vice versa, may be due to more available variance in the latter, relative to the former, at Time 2. The contemporaneous measurement of family stress processes and at only two time points, limits our ability to model cascading effects – as predicted by the FSM – properly over time. More measurement time points, combined with a true, autoregressive, cross-lagged model, in future research might provide insight into whether our observations are unique to our design or implicate a substantive extension of the FSM.

Our findings related to delay of gratification may be unique to the present sample and task paradigm. Black children from low-income families have shown unique sensitivities to aspects of social context, possibly through a mediating role of trust in experimenters (Price-Williams & Ramirez, 1974). Further, choice delay tasks are not related to EF (Duckworth & Kern, 2011), and may therefore not be influenced by family stress processes in the same way as other DoG tasks tapping EF-related abilities. Additionally, choice delay tasks require children to forego contact with *all* rewards during the hours waiting for larger-later rewards, and may carry more perceived risk compared to other tasks which do not involve extended loss of contact with research staff and rewards.

A natural extension of this study would be exploration of moderators of our observed effects. Examples include family characteristics and coping mechanisms which disrupt the pathways between economic hardship and family stress and/or build resilience in children.

Implications and Conclusion

To support children from low-income families, it is critical to understand the family processes that most strongly impact child development. A deep body of FSM research has elucidated family stress effects for a host of developmental outcomes in children (see Conger et al., 2010). This study broadens insights by examining contributions of family stress processes to two self-regulatory abilities implicated in academic success (Ursache et al., 2012), and improved health and wealth later in life (Moffit et al., 2011). Results indicate changes in family stress processes relate to improvement in these outcomes, and underscore the importance of the child-caregiver interactions as a key contributor to child outcomes (e.g., EF), and therefore a worthy target for intervention. But, results should also inspire future research to investigate the diverse pathways from family stress to children's development. The mediating role of child-caregiver interactions – a key aspect of the FSM – was not supported here for development of children's delay of gratification; instead, increase in financial strain was negatively related to increases in proclivity to delay gratification. This finding is novel, and caution is warranted in assuming its generalizability beyond the present study's target population and particular measure of delay of gratification. Nevertheless, fully alleviating poverty's effects on children may not be possible without addressing its more immediate consequences on caregivers.

References

Acock, A. C. (2005). Working with missing values. *Journal of Marriage and family*, *67*, 1012-1028. doi:10.1111/j.1741-3737.2005.00191.x

Anderson, R. E. (2015). Focusing on Family: Parent–Child Relationships and School Readiness among Economically Impoverished Black Children. *The Journal of Negro Education*, *84*, 442-456. doi:10.7709/jnegroeducation.84.3.0442

Best, J. R., Miller, P. H., & Jones, L. L. (2009). Executive functions after age 5: Changes and correlates. *Developmental Review*, *29*, 180-200. doi:10.1016/j.dr.2009.05.002

Blair, C., Granger, D. A., Willoughby, M., Mills-Koonce, R., Cox, M., Greenberg, M. T., ... & Fortunato, C. K. (2011). Salivary cortisol mediates effects of poverty and parenting on executive functions in early childhood. *Child development*, *82*, 1970-1984. doi:10.1111/j.1467-8624.2011.01643.x

Blair, C., & Raver, C. C. (2012). Child development in the context of adversity: experiential canalization of brain and behavior. *American Psychologist*, *67*, 309. doi:10.1037/a0027493

Blair, C., Raver, C. C., & Berry, D. J. (2014). Two approaches to estimating the effect of parenting on the development of executive function in early childhood. *Developmental Psychology*, *50*(2), 554. doi: 10.1037/a0033647

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child development*, *78*, 647-663. doi:10.1111/j.1467-8624.2007.01019.x

Bronfenbrenner, U. (1986). Ecology of the family as a context for human development: Research perspectives. Developmental psychology, 22(6), 723.

Cameron Ponitz, C., McClelland, M. M., Matthews, J. S., & Morrison, F. J. (2009). A structured observation of behavioral self-regulation and its contribution to kindergarten outcomes. *Developmental psychology*, 45, 605. doi:10.1037/a0015365

Cardoso, J. B., Padilla, Y. C., & Sampson, M. (2010). Racial and ethnic variation in the predictors of maternal parenting stress. *Journal of Social Service Research, 36,* doi:10.1080/01488376.2010.510948

Cassells, R. C., & Evans, G. W. (2017). Ethnic Variation in Poverty and Parenting Stress. In Deater-Deckar, K. & Panneton, R. (Eds.) *Parental Stress and Early Child Development: Adaptive and Maladaptive Outcomes.* Springer International Publishing AG, Switzerland

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of health and social behavior*, 385-396.

Conger, R. D., & Elder Jr, G. H. (1994). *Families in Troubled Times: Adapting to Change in Rural America. Social Institutions and Social Change*. Aldine de Gruyter, 200 Saw Mill River Road, Hawthorne, NY 10532

Conger, R. D., Conger, K. J., & Martin, M. J. (2010). Socioeconomic status, family processes, and individual development. *Journal of Marriage and Family*, *72*, 685-704. doi:10.1111/j.1741-3737.2010.00725.x

Conger, R. D., Wallace, L. E., Sun, Y., Simons, R. L., McLoyd, V. C., & Brody, G. H. (2002). Economic pressure in African American families: a replication and extension of the family stress model. *Developmental psychology*, *38*, 179. doi:10.1037/0012-1649.38.2.179

Costello, E. J., Compton, S. N., Keeler, G., & Angold, A. (2003). Relationships between poverty and psychopathology: A natural experiment. *Jama*, *290*, 2023-2029. doi:10.1001/jama.290.15.2023

Duckworth, A. L., & Kern, M. L. (2011). A meta-analysis of the convergent validity of self-control measures. *Journal of Research in Personality*, *45*(3), 259-268.

Duncan, G. J., Coe, R., Corcoran, M., Hill, M. S., Hoffman, S., & Morgan, J. N. (1984). *Years of poverty--years of plenty: The changing economic fortunes of American workers and families*.

Evans, G. W. (2003). A multimethodological analysis of cumulative risk and allostatic load among rural children. *Developmental psychology*, *39*, 924. doi:10.1037/0012-1649.39.5.924

Evans, G. W. (2004). The environment of childhood poverty. American psychologist, 59(2), 77.Evans, G. W., & English, K. (2002). The environment of poverty: Multiple stressor exposure, psychophysiological stress, and socioemotional adjustment. *Child development*, *73*, 1238-1248. doi:10.1037/0003-066X.59.2.77

Evans, G. W., Gonnella, C., Marcynyszyn, L. A., Gentile, L., & Salpekar, N. (2005). The role of chaos in poverty and children's socioemotional adjustment. *Psychological science*, *16*, 560-565. doi:10.1111/j.0956-7976.2005.01575.x

Evans, G. W., & Kim, P. (2013). Childhood Poverty, Chronic Stress, Self-Regulation, and Coping. *Child Development Perspectives*, *7*, 43-48. doi:10.1111/cdep.12013

Evans, G. W., & Rosenbaum, J. (2008). Self-regulation and the income-achievement gap. *Early Childhood Research Quarterly*, *23*, 504-514. doi:10.1016/j.ecresq.2008.07.002

Evans, G. W., & Schamberg, M. A. (2009). Childhood poverty, chronic stress, and adult working memory. *Proceedings of the National Academy of Sciences*, *106*, 6545-6549. doi:10.1073/pnas.0811910106

Fawcett, T. W., McNamara, J. M., & Houston, A. I. (2012). When is it adaptive to be patient? A general framework for evaluating delayed rewards. *Behavioural processes*, *89*, 128-136. doi:10.1016/j.beproc.2011.08.015

Finegood, E. D. & Blair, C. (2017). Poverty, Parent Stress, and Emerging Executive Functions in Young Children. In Deater-Deckar, K. & Panneton, R. (Eds.) *Parental Stress and Early Child Development: Adaptive and Maladaptive Outcomes.* Springer International Publishing AG, Switzerland

Hackman, D. A., Gallop, R., Evans, G. W., & Farah, M. J. (2015). Socioeconomic status and executive function: developmental trajectories and mediation. *Developmental Science*, *18*, 686-702. doi:10.1111/desc.12246

Hongwanishkul, D., Happaney, K. R., Lee, W. S., & Zelazo, P. D. (2005). Assessment of hot and cool executive function in young children: Age-related changes and individual differences. *Developmental neuropsychology*, *28*, 617-644. doi:10.1207/s15326942dn2802_4

Houck, G. M. & Lecuyer-Maus, E. A. (2004). Maternal limit setting during toddlerhood, delay of gratification, and behavior problems at age five. *Infant Mental Health Journal*, *25*, 28-46. DOI: 10.1002/imhj.10083

Hu, L. T. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, *6*, 1-55. doi:10.1080/10705519909540118

Jacobs, J. E. & Klaczynski, P. A. (2002). The development of judgment and decision making during childhood and adolescence. Current directions in psychological science, 11(4), 145-149.

Kidd, C., Palmeri, H., & Aslin, R. N. (2013). Rational snacking: Young children's decision-making on the marshmallow task is moderated by beliefs about environmental reliability. *Cognition, 126*, 109-114. doi:10.1016/j.cognition.2012.08.004

Lee, W. S., & Carlson, S. M. (2015). Knowing when to be “rational”: Flexible economic decision making and executive function in preschool children. *Child Development*, *86*(5), 1434-1448. doi: 10.1111/cdev.12401

Lengua, L. J., Honorado, E., & Bush, N. R. (2007). Contextual risk and parenting as predictors of effortful control and social competence in preschool children. *Journal of applied developmental psychology, 28*, 40-55. doi:10.1016/j.appdev.2006.10.001

Lengua, L. J., Moran, L., Zalewski, M., Ruberry, E., Kiff, C., & Thompson, S. (2015). Relations of growth in effortful control to family income, cumulative risk, and adjustment in preschool-age children. *Journal of abnormal child psychology*, *43*, 705-720. doi:10.1007/s10802-014-9941-2

Lengua, L. J., Zalewski, M., Fisher, P., & Moran, L. (2013). Does HPA-Axis Dysregulation Account for the Effects of Income on Effortful Control and Adjustment in Preschool Children? *Infant and child development*, *22*, 439-458. doi:10.1002/icd.1805MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research, 39*(1), 99. doi:10.1207/s15327906mbr3901_4

McClelland, M. M., Cameron, C. E., Duncan, R., Bowles, R. P., Acock, A. C., Miao, A., & Pratt, M. E. (2014). Predictors of early growth in academic achievement: the head-toes-knees-shoulders task. *Frontiers in Psychology*, *5*, 599.

McLoyd, V. C. (1998). Socioeconomic disadvantage and child development. *American psychologist*, *53*, 185. doi:10.1037/0003-066X.53.2.185

Mischel, W., & Ayduk, O. (2011). Willpower in a cognitive-affective processing system. In Vohs, K. D., & Baumeister, R. F. (Eds.). (2011). *Handbook of self-regulation: Research, theory, and applications*. Guilford Press. 83-105.

Mistry, R. S., Biesanz, J. C., Taylor, L. C., Burchinal, M., & Cox, M. J. (2004). Family income and its relation to preschool children's adjustment for families in the NICHD Study of Early Child Care. *Developmental psychology, 40*, 727-745. doi: 10.1037/0012-1649.40.5.727

Mistry, R. S., Vandewater, E. A., Huston, A. C., & McLoyd, V. C. (2002). Economic Well-Being and Children's Social Adjustment: The Role of Family Process in an Ethnically Diverse Low-Income Sample. *Child development*, *73*, 935-951. doi:10.1111/1467-8624.00448

Mistry, R. S., & Wadsworth, M. (2011). Family Functioning and Child Development in the Context of Poverty. *Prevention Researcher*, *18*, 11-15.

Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H. L., . . . Ross, S. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences, 108*, 2693-2698. doi:10.1073/pnas.1010076108

Muthén, L. K., & Muthén, B. O. (2005). *Mplus: Statistical analysis with latent variables:*

User's guide (pp. 1998-2012). Los Angeles: Muthén & Muthén.Neville, H. J., Stevens, C., Pakulak, E., Bell, T. A., Fanning, J., Klein, S., & Isbell, E. (2013). Family-based training program improves brain function, cognition, and behavior in lower socioeconomic status preschoolers. *Proceedings of the National Academy of Sciences*, *110*, 12138-12143. doi:10.1073/pnas.1304437110

Noble, K. G., Norman, M. F., & Farah, M. J. (2005). Neurocognitive correlates of socioeconomic status in kindergarten children. *Developmental science*, *8*, 74-87. doi:10.1111/j.1467-7687.2005.00394.x

Parke, R. D., Coltrane, S., Duffy, S., Buriel, R., Dennis, J., Powers, J., ... & Widaman, K. F. (2004). Economic stress, parenting, and child adjustment in Mexican American and European American families. *Child development*, *75*(6), 1632-1656.

Pianta, R. C. (1992). Child-parent relationship scale. *Unpublished measure*, *University of Virginia*.

Price-Williams, D. R., & Ramirez III, M. (1974). Ethnic differences in delay of gratification. *The Journal of Social Psychology*, *93*(1), 23-30.

Raver, C. C., Blair, C., Willoughby, M., & The Family Life Project Key Investigators. (2013). Poverty as a predictor of 4-year-olds' executive function: New perspectives on models of differential susceptibility. *Developmental Psychology*, *49*(2), 292-304.

Raver, C. C., Gershoff, E. T., & Aber, J. L. (2007). Testing equivalence of mediating models of income, parenting, and school readiness for White, Black, and Hispanic children in a national sample. *Child development*, *78*(1), 96-115. doi: 10.1111/j.1467-8624.2007.00987.x

Rosati, Alexandra G. and Stevens, Jeffrey R., Rational Decisions: The Adaptive Nature of Context-Dependent Choice (2009). Faculty Publications, Department of Psychology. Paper 525. http://digitalcommons.unl.edu/psychfacpub/525

Ross, C. E. & Van Willigen, M. (1997). Education and the subjective quality of life. *Journal of Health and Social Behavior*, *38*(3), 275-297.

Strickland, B. R. (1972). Delay of gratification as a function of race of the experimenter. *Journal of personality and Social Psychology*, *22*(1), 108.

Ursache, A., Blair, C., & Raver, C. C. (2012). The promotion of self-regulation as a means of enhancing school readiness and early achievement in children at risk for school failure. *Child Development Perspectives*, *6*, 122-128. doi:10.1111/j.1750-8606.2011.00209.x

Vinokur, A. D., Price, R. H., & Caplan, R. D. (1996). Hard times and hurtful partners: how financial strain affects depression and relationship satisfaction of unemployed persons and their spouses. *Journal of personality and social psychology*, *71*, 166. doi:10.1037/0022-3514.71.1.166

Wadsworth, M. E. (2015). Development of maladaptive coping: A functional adaptation to chronic, uncontrollable stress. *Child development perspectives*, *9*(2), 96-100.

Yeung, W. J., Linver, M. R., & Brooks–Gunn, J. (2002). How money matters for young children's development: Parental investment and family processes. *Child development*, *73*(6), 1861-1879.

Tables

Table 1. Univariate Descriptives and Bivariate Correlations Among Study Variables

		n	Missing (%)	Mean	SD	Min	Max	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Black	301	12	0.89	0.32	0	1																
	Latino	301	12	0.06	0.24	0	1																
	FRL	296	14	0.96	0.21	0	1																
1.	Age T1	328	5	5.43	0.33	4.5	6.17																
2.	Treatment	343	0	0.6	0.49	0	1	-.03															
3.	Male	343	0	0.46	0.5	0	1	.00	-.04														
4.	Cohort 2	343	0	0.29	0.45	0	1	.03	-.03	.01													
5.	Cohort 3	343	0	0.31	0.46	0	1	.08	.00	.04	-.43												
6.	HS Drop	300	13	0.29	0.45	0	1	.05	-.07	.03	-.03	.08											
7.	Pub Assist	302	12	0.81	0.39	0	1	.04	.02	-.10	.02	-.02	.10										
8.	FS T1	302	12	2.02	0.91	1	5	.05	.05	-.03	.00	-.01	-.08	.10									
9.	GS T1	302	12	2.56	0.57	1.07	4.21	.02	.00	.07	-.06	-.04	.08	.13	.50								
10.	Con T1	302	12	1.89	0.77	1	4.38	-.12	.01	.04	-.06	.08	.07	-.07	.08	.29							
11.	HTKS T1	328	4	28.61	27.2	0	88	.14	-.13	-.18	.10	.02	-.11	.01	-.06	-.08	-.06						
12.	CD T1	328	4	3.16	2.2	0	6	-.03	.05	-.05	.03	-.10	.02	-.05	.06	-.01	-.02	-.20					
13.	FS T2	238	31	1.89	0.87	1	5	-.09	-.09	.10	-.01	-.08	.04	.14	.31	.35	.16	-.11	-.06				
14.	GS T2	238	31	2.53	0.59	1.29	4.43	-.09	.01	.10	.01	-.06	.08	.11	.20	.51	.36	-.12	-.04	.46			
15.	Con T2	239	30	1.9	0.73	1	4.12	-.02	.00	.10	-.04	.02	.14	-.07	.11	.22	.58	-.14	.07	.18	.46		
16.	HTKS T2	268	22	53.75	26.68	0	59	.06	.03	-.16	.05	-.14	-.11	-.03	.02	-.11	-.17	.50	-.12	-.10	-.15	-.30	
17.	CD T2	275	20	3.56	2.22	0	6	.05	.07	-.04	-.03	-.10	.06	.04	-.11	-.08	-.06	.01	.30	-.18	-.06	.05	.01

Notes: T1 = Time 1; T2 = Time 2; Treatment Coded as 1 = Offered Treatment; HS Drop = Mother High School Dropout; Pub Assist = Family Receiving Public Assistance; FS = Caregiver Financial Strain; GS = Caregiver General Stress; Con = Child-Caregiver Conflict; HTKS = Head-Toes-Knees-Shoulders; CD = Choice Delay. **Bold** indicates statistically significant correlations (p < .05). Means for binary variables (Treatment, Male, Cohort 2, Cohort 3, HS Drop, Public Assistance) represent percentages.

Table 2. Path Model Direct Effects on Child Outcomes

	HTKS					Choice Delay				
	Est.	SE	Std. Est.	p-value		Est.	SE	Std. Est.	p-value	
Child Outcome T1 ON										
Mother HS Dropout	-7.81	3.39	-0.13	0.02	*	0.26	0.29	0.05	0.36	
Public Assistance	0.42	3.83	0.01	0.91		-0.38	0.35	-0.07	0.29	
CG Financial Strain T1	-2.03	1.91	-0.07	0.29		0.20	0.16	0.08	0.23	
CG General Stress T1	-0.49	3.43	-0.01	0.89		-0.18	0.27	-0.05	0.51	
Child-CG Conflict T1	-0.46	2.07	-0.01	0.83		-0.10	0.18	-0.04	0.57	
Child Outcome T2 ON										
Mother HS Dropout	1.37	3.45	0.02	0.69		0.16	0.30	0.03	0.58	
Public Assistance	-3.49	4.00	-0.05	0.38		0.53	0.34	0.09	0.11	
CG Financial Strain T1	3.31	2.04	0.11	0.11		-0.21	0.20	-0.09	0.29	
CG General Stress T1	-3.90	3.57	-0.08	0.28		-0.15	0.33	-0.04	0.65	
Child-CG Conflict T1	0.26	2.60	0.01	0.92		-0.24	0.24	-0.08	0.31	
CG Financial Strain T2	-0.63	2.42	-0.02	0.80		-0.41	0.20	-0.16	0.04	*
CG General Stress T2	1.20	3.96	0.03	0.76		0.09	0.33	0.02	0.80	
Child-CG Conflict T2	-8.52	2.88	-0.23	0.00	**	0.42	0.24	0.14	0.08	
Child Outcome T1	0.48	0.06	0.49	0.00	***	0.28	0.06	0.28	0.00	***
R^2 in Child Outcome T1	0.10					0.03				
R^2 in Child Outcome T2	0.35					0.17				

Notes: T1 = Time 1; T2 = Time 2; CG = Caregiver; HS = High School.

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Table 3. Total, Total Indirect, and Specific Indirect Effects on Time 2 HTKS

	Estimate	Lower 2.5%	Lower 5%	Upper 5%	Upper 2.5%
HS Dropout					
Total Effect	-0.087	-0.214	-0.194	0.020	0.041
Total Indirect Effect	-0.110	-0.179	-0.168	-0.052	-0.041
Specific Indirect Effects (significant only):					
HSDrop-HTKS1-HTKS2	-0.064	-0.122	-0.112	-0.016	-0.007
Public Assistance					
Total Effect	-0.043	-0.173	-0.152	0.066	0.087
Total Indirect Effect	0.008	-0.060	-0.049	0.066	0.077
Specific Indirect Effects (none significant)					
CG Financial Strain Time 1					
Total Effect	0.000	-0.167	-0.127	0.107	0.127
Total Indirect Effect	-0.112	-0.211	-0.195	-0.029	-0.014
Specific Indirect Effects (significant only):					
FS1-GS1-Con1-Con2-HTKS2	-0.019	-0.035	-0.033	-0.006	-0.003
FS1-GS1-GS2-Con2-HTKS2	-0.015	-0.030	-0.027	-0.002	0.000
CG General Stress Time 1					
Total Effect	-0.123	-0.273	-0.249	0.022	0.026
Total Indirect Effect	-0.040	-0.151	-0.133	0.053	0.071
Specific Indirect Effects (significant only):					
GS1-Con1-Con2-HTKS2	-0.038	-0.069	-0.064	-0.012	-0.007
GS1-GS2-Con2-HTKS2	-0.029	-0.057	-0.053	-0.006	-0.001
Child-CG Conflict Time 1					
Total Effect	-0.123	-0.260	-0.238	-0.009	0.013
Total Indirect Effect	-0.131	-0.238	-0.221	-0.041	-0.024
Specific Indirect Effects (significant only):					
Con1-Con2-HTKS2	-0.112	-0.194	-0.181	-0.043	-0.029
Con1-GS2-Con2-HTKS2	-0.016	-0.031	-0.029	-0.004	-0.001
CG Financial Strain T2					
Total Effect	-0.030	-0.165	-0.144	0.083	0.105
Total Indirect Effect	-0.010	-0.075	-0.064	0.045	0.055
Specific Indirect Effects (significant only):					
FS2-GS2-Con2-HTKS2	-0.025	-0.051	-0.047	-0.004	0.000
CG General Stress T2					
Total Effect	-0.052	-0.227	-0.199	0.094	0.122
Total Indirect Effect	-0.079	-0.145	-0.134	-0.024	-0.014
Specific Indirect Effects (significant only):					
GS2-Con2-HTKS2	-0.079	-0.145	-0.134	-0.024	-0.014

Notes: Only significant specific indirect effects shown. Confidence intervals are estimated based on bootstrapped confidence intervals. HS = High School; FS1 = Time 1 Financial Strain; GS1 = Time 1 General Stress; Con1 = Time 1 Child-Caregiver Conflict; HTKS2 = HTKS Time 2; FS2 = Time 2 Financial Strain; GS2 = Time 2 General Stress; Con2 = Time 2 Child-Caregiver Conflict; HTKS2 = HTKS Time 2.

Supplemental Tables

Table S4. Parameter Estimates for HTKS and Choice Delay Path Models

	HTKS				Choice Delay			
	Est.	S.E.	β	p-value	Est.	S.E.	Β	p-value
CG Financial Strain T1 ON								
Mother HS Dropout	-0.17	0.11	-0.09	0.12	-0.17	0.11	-0.09	0.12
Public Assistance	0.24	0.14	0.10	0.09	0.24	0.14	0.10	0.08
Male	-0.02	0.11	-0.03	0.82	-0.02	0.11	-0.02	0.84
Offered Treatment	0.09	0.11	0.10	0.41	0.08	0.11	0.09	0.45
Age T1	0.17	0.18	0.06	0.33	0.17	0.18	0.06	0.34
Cohort 2	-0.02	0.12	-0.02	0.87	-0.03	0.12	-0.03	0.82
Cohort 3	-0.02	0.12	-0.02	0.86	-0.03	0.12	-0.04	0.78
CG General Stress T1 ON								
Mother HS Dropout	0.14	0.07	0.11	0.03 *	0.14	0.07	0.11	0.03 *
Public Assistance	0.11	0.08	0.08	0.14	0.11	0.08	0.08	0.14
CG Financial Strain T1	0.32	0.03	0.51	0.00 ***	0.32	0.03	0.50	0.00 ***
Male	0.11	0.06	0.19	0.06	0.11	0.06	0.20	0.05
Offered Treatment	-0.03	0.06	-0.04	0.67	-0.02	0.06	-0.04	0.70
Age T1	-0.06	0.09	-0.03	0.54	-0.02	0.09	-0.01	0.87
Cohort 2	-0.12	0.07	-0.20	0.09	-0.12	0.07	-0.21	0.09
Cohort 3	-0.09	0.07	-0.15	0.23	-0.09	0.07	-0.15	0.22
Child-CG Conflict T1 ON								
Mother HS Dropout	0.07	0.10	0.04	0.47	0.07	0.10	0.04	0.49
Public Assistance	-0.21	0.12	-0.11	0.09	-0.20	0.12	-0.10	0.10
CG Financial Strain T1	-0.06	0.06	-0.07	0.30	-0.06	0.05	-0.07	0.27
CG General Stress T1	0.46	0.08	0.34	0.00 ***	0.46	0.08	0.35	0.00 ***
Male	-0.01	0.09	-0.02	0.89	-0.01	0.09	-0.02	0.90
Offered Treatment	0.02	0.09	0.02	0.84	0.02	0.09	0.03	0.80
Age T1	-0.33	0.14	-0.14	0.02 *	-0.31	0.14	-0.13	0.02 *
Cohort 2	0.03	0.10	0.04	0.78	0.03	0.10	0.04	0.77
Cohort 3	0.18	0.11	0.23	0.09	0.18	0.11	0.24	0.08
Child Outcome T1 ON								
Mother HS Dropout	-7.81	3.39	-0.13	0.02 *	0.26	0.29	0.05	0.36
Public Assistance	0.42	3.83	0.01	0.91	-0.38	0.35	-0.07	0.29
CG Financial Strain T1	-2.03	1.91	-0.07	0.29	0.20	0.16	0.08	0.23
CG General Stress T1	-0.49	3.43	-0.01	0.89	-0.18	0.27	-0.05	0.51
Child-CG Conflict T1	-0.46	2.07	-0.01	0.83	-0.10	0.18	-0.04	0.57
Male	-9.90	2.90	-0.36	0.00 **	-0.20	0.26	-0.09	0.43
Offered Treatment	-7.68	3.10	-0.28	0.01 *	0.25	0.24	0.11	0.31
Age T1	11.48	4.56	0.14	0.01 *	-0.17	0.39	-0.03	0.66
Cohort 2	7.53	3.48	0.28	0.03 *	-0.06	0.30	-0.03	0.84
Cohort 3	4.71	3.61	0.17	0.19	-0.49	0.30	-0.22	0.10
CG Financial Strain T2 ON								
Mother HS Dropout	0.09	0.13	0.05	0.49	0.11	0.12	0.06	0.38
Public Assistance	0.26	0.13	0.12	0.05 *	0.25	0.13	0.11	0.06
CG Financial Strain T1	0.24	0.08	0.25	0.00 **	0.25	0.08	0.25	0.00 **
CG General Stress T1	0.28	0.12	0.18	0.02 *	0.28	0.12	0.18	0.02 *

Child-CG Conflict T1	0.07	0.08	0.06	0.39		0.07	0.08	0.06	0.38	
Child Outcome T1	0.00	0.00	-0.04	0.47		-0.03	0.02	-0.08	0.16	
Male	0.18	0.11	0.21	0.10		0.19	0.11	0.21	0.09	
Offered Treatment	-0.13	0.12	-0.15	0.28		-0.11	0.12	-0.12	0.37	
Age T1	-0.20	0.18	-0.07	0.28		-0.23	0.18	-0.08	0.20	
Cohort 2	-0.05	0.14	-0.06	0.71		-0.06	0.13	-0.07	0.66	
Cohort 3	-0.14	0.13	-0.17	0.25		-0.15	0.12	-0.17	0.23	
CG General Stress T2 ON										
Mother HS Dropout	0.06	0.07	0.05	0.40		0.07	0.07	0.05	0.34	
Public Assistance	0.04	0.08	0.02	0.66		0.03	0.08	0.02	0.73	
CG Financial Strain T1	-0.05	0.04	-0.07	0.27		-0.05	0.04	-0.07	0.30	
CG General Stress T1	0.38	0.07	0.37	0.00	***	0.39	0.07	0.37	0.00	***
Child-CG Conflict T1	0.16	0.05	0.20	0.00	**	0.16	0.05	0.21	0.00	***
CG Financial Strain T2	0.22	0.05	0.32	0.00	***	0.22	0.05	0.32	0.00	***
Child Outcome T1	0.00	0.00	-0.03	0.59		-0.01	0.01	-0.04	0.51	
Male	0.05	0.07	0.09	0.40		0.06	0.06	0.10	0.36	
Offered Treatment	0.10	0.06	0.17	0.10		0.11	0.06	0.19	0.09	
Age T1	-0.03	0.10	-0.02	0.78		-0.04	0.10	-0.02	0.67	
Cohort 2	0.09	0.08	0.16	0.23		0.09	0.08	0.16	0.23	
Cohort 3	-0.04	0.08	-0.07	0.57		-0.05	0.08	-0.09	0.52	
Child-CG Conflict T2 ON										
Mother HS Dropout	0.14	0.09	0.09	0.13		0.14	0.10	0.09	0.14	
Public Assistance	-0.05	0.12	-0.03	0.65		-0.06	0.12	-0.03	0.60	
CG Financial Strain T1	0.07	0.05	0.09	0.13		0.06	0.05	0.08	0.17	
CG General Stress T1	-0.18	0.09	-0.14	0.06		-0.18	0.09	-0.14	0.06	
Child-CG Conflict T1	0.46	0.07	0.48	0.00	***	0.46	0.06	0.48	0.00	***
CG Financial Strain T2	-0.03	0.05	-0.03	0.62		-0.02	0.05	-0.02	0.70	
CG General Stress T2	0.42	0.10	0.34	0.00	***	0.43	0.10	0.35	0.00	***
Child Outcome T1	0.00	0.00	-0.07	0.16		0.02	0.02	0.07	0.25	
Male	0.11	0.08	0.15	0.15		0.12	0.08	0.16	0.13	
Offered Treatment	-0.01	0.08	-0.02	0.90		-0.01	0.08	-0.01	0.94	
Age T1	0.19	0.13	0.09	0.13		0.17	0.12	0.08	0.17	
Cohort 2	0.00	0.09	0.00	0.98		-0.01	0.09	-0.02	0.90	
Cohort 3	-0.01	0.09	-0.01	0.93		-0.02	0.10	-0.02	0.87	
Child Outcome T2 ON										
Mother HS Dropout	1.37	3.45	0.02	0.69		0.16	0.30	0.03	0.58	
Public Assistance	-3.49	4.00	-0.05	0.38		0.53	0.34	0.09	0.11	
CG Financial Strain T1	3.31	2.04	0.11	0.11		-0.21	0.20	-0.09	0.29	
CG General Stress T1	-3.90	3.57	-0.08	0.28		-0.15	0.33	-0.04	0.65	
Child-CG Conflict T1	0.26	2.60	0.01	0.92		-0.24	0.24	-0.08	0.31	
CG Financial Strain T2	-0.63	2.42	-0.02	0.80		-0.41	0.20	-0.16	0.04	*
CG General Stress T2	1.20	3.96	0.03	0.76		0.09	0.33	0.02	0.80	
Child-CG Conflict T2	-8.52	2.88	-0.23	0.00	**	0.42	0.24	0.14	0.08	
Child Outcome T1	0.48	0.06	0.49	0.00	***	0.28	0.06	0.28	0.00	***
Male	-1.70	2.85	-0.06	0.55		-0.08	0.27	-0.04	0.77	
Offered Treatment	4.86	2.82	0.18	0.09		0.19	0.27	0.08	0.48	
Age T1	-0.53	4.62	-0.01	0.91		0.23	0.44	0.03	0.61	
Cohort 2	-5.10	3.45	-0.19	0.14		-0.58	0.32	-0.26	0.07	
Cohort 3	-10.55	3.49	-0.40	0.00	**	-0.72	0.32	-0.32	0.03	*

Mother HS Dropout WITH

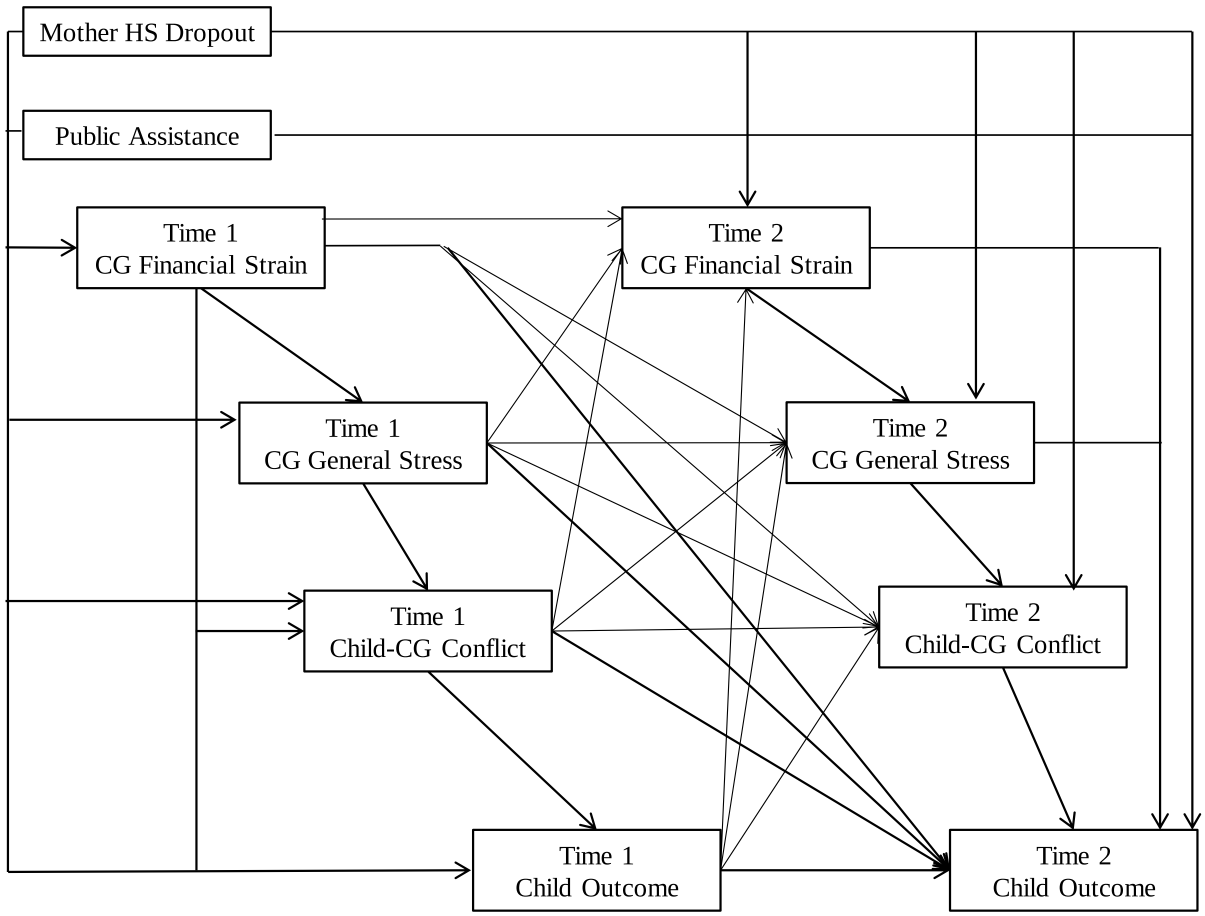
Age T1	0.01	0.01	0.05	0.42		0.01	0.01	0.05	0.44	
Public Assistance WITH										
Age T1	0.01	0.01	0.04	0.55		0.01	0.01	0.04	0.51	
Mother HS Dropout	0.02	0.01	0.10	0.06		0.02	0.01	0.10	0.06	
Variances										
Age T1	0.11	0.01	1.00	0.00	***	0.11	0.01	1.00	0.00	***
Mother HS Dropout	0.21	0.01	1.00	0.00	***	0.21	0.01	1.00	0.00	***
Public Assistance	0.15	0.01	1.00	0.00	***	0.15	0.01	1.00	0.00	***
Residual Variances										
Child Outcome T1	663.51	38.14	0.90	0.00	***	4.71	0.21	0.97	0.00	***
Child Outcome T2	460.20	38.66	0.65	0.00	***	4.08	0.26	0.83	0.00	***
Child-CG Conflict T1	0.51	0.04	0.87	0.00	***	0.51	0.04	0.87	0.00	***
Child-CG Conflict T2	0.30	0.03	0.57	0.00	***	0.30	0.03	0.57	0.00	***
CG Financial Strain T1	0.80	0.07	0.98	0.00	***	0.80	0.07	0.98	0.00	***
CG Financial Strain T2	0.60	0.06	0.79	0.00	***	0.60	0.06	0.78	0.00	***
CG General Stress T1	0.23	0.02	0.71	0.00	***	0.23	0.02	0.71	0.00	***
CG General Stress T2	0.20	0.02	0.58	0.00	***	0.20	0.02	0.57	0.00	***
R-SQUARE										
Child Outcome T1	0.10					0.03				
Child Outcome T2	0.35					0.17				
CG Financial Strain T1	0.13					0.02				
CG Financial Strain T2	0.44					0.22				
CG General Stress T1	0.02					0.29				
CG General Stress T2	0.21					0.43				
Child-CG Conflict T1	0.29					0.13				
Child-CG Conflict T2	0.43					0.43				

Table 5. Chi-Squared Difference Tests Comparing Constrained vs Unconstrained Multi-group Models (Sensitivity Analysis Testing for Interference of Intervention with Results)

				Chi-Square Difference Tests		
	χ^2 Value	df	*p*-value	$\Delta\chi^2$ Value	Δdf	*p*-value
HTKS						
Unconstrained Model	22.82	18	0.20	71.69	76	0.62
Constrained Model	92.227	94	0.5324			
Choice Delay						
Unconstrained Model	24.11	18	0.15	74.40	76	0.53
Constrained Model	98.51	94	0.35			

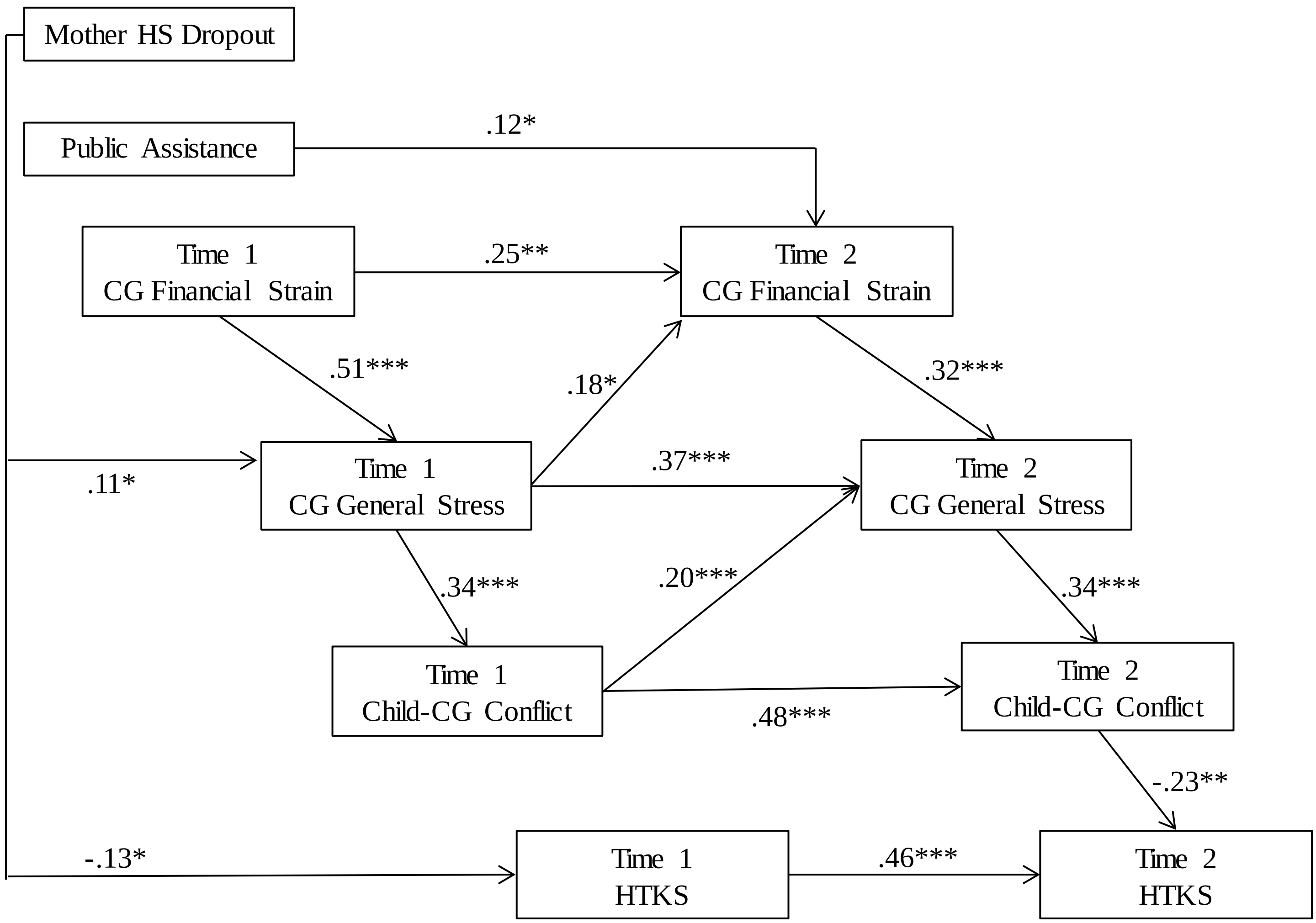
Figures

Figure 1. Path Model Specification



NOTES: TWO separate path models run, one for each outcome (HTKS and Choice Delay). CG = Caregiver; HS = High School. All paths shown in this figure were unconstrained; solid paths indicate paths predicted by the FSM. Covariates were: Gender (male); treatment condition; child age at Time 1.

Figure 2. Path Model Results with HTKS as Child Outcome



NOTES: CG = Caregiver; HS = High School. Covariates were: Gender (male); treatment condition; child age at Time 1.
*$p < .05$, **$p < .01$, ***$p < .001$.

PAPER 3: Validation of a Choice Delay of Gratification Measure among Children of Color from Low-Income Families

Chelsea A. K. Duran

Acknowledgements:

The authors sincerely thank the schools, teachers, and families who participated in this research and without whom this study would not have been possible.

The research reported here was supported by awards from by the Institute of Education Sciences, U.S. Department of Education, through Grant #R305A110703; by the Social Innovation Fund of the Corporation for National and Community Service, managed privately by the Edna McConnell Clark Foundation; and by the Institute of Education Sciences, U.S. Department of Education, through Grant #R305B090002 to the University of Virginia. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Validation of a Choice Delay of Gratification Measure among Children of Color from Low-Income Families

Successfully accumulating wealth, maintaining health, and attaining high levels of education are ostensibly predicated on choosing and pursuing long-term goals or well-being and, often correspondingly, abstaining from more immediately gratifying experiences. A tendency to delay gratification in childhood has long been understood to relate to well-being in many domains of life (Duckworth, 2011; Mischel, 2014), including in adulthood (Casey et al., 2011; Moffit et al., 2011). Whether these relations hold for all delay of gratification tasks, in particular those measuring only delay-related choices, and not the sustenance of one's self-imposed choice, is unclear (Bowersox, 2013; Hongwanishkul, Happaney, Lee, & Zelazo, 2005; Prencipe & Zelazo, 2005; Spann, 2014; Tynan, 2014). Further, few validation studies have been carried out among children from low-income families and/or of racial/ethnic minority status. This is potentially problematic since preferences toward immediate gratification may be reflective of adaptive development in the context of poverty (Ellis, Bianchi, Griskevicius, & Frankenhuis, 2017; Lee & Carlson, 2015; Sturge-Apple et al., 2016) and processes underlying delay choices may vary by race (Price-Williams & Ramirez, 1974; Strickland, 1972). And so, investigation into choice delay relations with other child outcomes in under-represented groups is pertinent.

To fill this gap in understanding, the present study examines the psychometric properties of a choice delay of gratification task among 372 children of color from low income families (average age = 5.5 years). Specifically, using secondary data from a longitudinal evaluation of a social and emotional learning (SEL) program, I first investigate the internal consistency and coherence of children's choice delay of gratification performance through item factor analysis (IFA). Second, I examine the extent to which choice delay performance corresponds to other, concurrently measured self-regulatory constructs. Third, I examine the predictive validity of choice delay performance in kindergarten for school-related outcomes in the first two years of elementary school. Results from this study have implications for the use of choice delay measures in empirical research, the refinement of emerging theories relating to delay of gratification as a functional adaptation, and the implementation of programs targeting non-academic skills in children of color from low-income families.

Traditional and Emerging Theoretical Accounts of Delay of Gratification Performance

The most popular theory accounting for delay of gratification and its predictive power assumes greater tendency toward delaying gratification indicates greater self-control, or a triumph of rational, cognitive processes over reflexive, emotional processes (Mischel & Ayduk, 2011). Delay of gratification tasks often offer children an option of consuming or partaking in a small, immediate reward or a larger reward available later on. Given this paper's focus on low-SES children, whose environments are defined, in part, by scarcity of physical resources, the use of concrete rewards as contingencies in these tasks is assumed to be an important design feature.

There are two primary types of delay of gratification tasks relevant to the present study's conceptualization of delay of gratification: sustained and choice delay. The well-known "marshmallow test" is a sustained delay task in which a researcher offers the child a choice between an immediate treat (e.g., a marshmallow) or a larger amount/number of said treat when the researcher returns an unspecified amount of time later (usually 15-30 minutes; Mischel & Ebbesen, 1970). The child is told he or she can eat the "smaller-sooner" treat at any time as he or she waits for the researcher to return with the "larger-later" treat. The present study is focused on a choice delay task, in which researchers offer a child a series of paired choices between smaller-sooner and larger-later rewards (e.g., stickers or candy). In a choice delay task, smaller-sooner rewards, when chosen, are given immediately and prior to the next choice(s). Larger-later rewards are presented alongside the smaller-sooner rewards but are typically furnished at least several hours later (up to weeks later, depending on the study) without the option during the waiting time to recant one's decision. Thus, sustained delay tasks theoretically measure both of two delay of gratification processes – first, detection and selection of the more desirable reward and, second, sustaining one's self-imposed choice – whereas choice delay tasks measure only the first (Fishbach & Converse, 2011; Mischel, 1974; Mischel, Shoda, & Rodriguez, 1989; Reynolds & Schiffbauer, 2005).

The bulk of the present literature focused on delaying gratification – and choice delay of gratification in particular – was based on samples of predominantly White children from families of middle to upper socioeconomic status (SES). Theoretically, poverty is a powerful influence on self-regulatory development, including the extent to which specific self-regulatory constructs reflect adaptive versus maladaptive development (Blair & Raver, 2012; Harris et al., 2017). This notion may apply in the case of delay of gratification in particular. A sequence of two studies found negative relations between vagal tone – a physiological index of effective self-regulation – in toddlerhood and sustained delay of gratification at age four among children of low socioeconomic status. The reverse was observed for middle-SES children (Sturge-Apple et al., 2016). That vagal tone positively related to delay of gratification in middle-SES children is consistent with the general consensus historically dominating the literature: that delay of gratification reflects strong self-regulatory ability. The observation of the *negative* association in lower-SES children calls into question whether this should be presumed for all children in all developmental contexts. Hence, the importance of considering whether delay of gratification choices are related to other aspects of children's functioning in a low-income sample.

Race and/or minority status may also represent an aspect of social context which carries unique influence(s) on children's delay of gratification behaviors. Among children from low-income families, Black sixth graders', but not White sixth graders', delay of gratification choices were sensitive to race of the experimenter administering the assessment, which was randomly assigned (Strickland, 1972). Perhaps relatedly, elementary-aged Black children from low-income families were observed to express trust in experimenters' promises of larger-later rewards less often compared to White children in the same low-income sample (Price-Williams & Ramirez, 1974). The authors of these studies speculate that Black children may be less trusting of White adults as a result of experiencing discrimination and prejudice (Price-Williams & Ramirez, 1974; Strickland, 1972). This speculation is consistent with experimental confirmation of perceived trustworthiness of individuals offering rewards as an influence on delay of gratification behaviors in both adults and young children (Michaelson et al., 2013; Michaelson & Munakata, 2016). Thus, the influence of an adaptive decision-making process makes the properties of choice delay tasks an interesting subject for study among Black children of low-SES.

Measurement Properties of Choice Delay Measures

Early in the history of delay of gratification research, in the 1950s and 1960s, sustained and choice delay measures were often considered conceptually similar measures. Since the 1970s, however, there have been numerous studies suggesting that the two tasks measuring distinct, although related, processes. First, there is mixed evidence for a developmental gradient in early childhood on choice delay tasks (Hongwanishkul et al., 2005; Thompson, Barresi, & Moore, 1997; Prencipe & Zelazo, 2005; Lindstrom & Shipman, 1972; Schwarz et al., 1983; Toner, Holstein, & Hetherington, 1977) and more consistent evidence for improvements with age in sustained delay tasks (see Mischel, Shoda, & Rodriguez, 1989). Ostensibly, these reflect self-regulatory abilities, including executive functions (Mischel, Ebbesen, & Zeiss, 1972), which have well-established developmental gradients during early childhood (Anderson, 2002). Second, there have been rare observations of negative correlations in some subsamples between performance on each type of task (Toner et al., 1977). Finally, as discussed further later, there are also differences in the extent to which these measures relate to other self-regulatory measures, as well as to academic outcomes. The present section provides an overview of psychometric properties of choice delay measures, as found in the literature. Where appropriate, comparisons are made with the same properties of sustained delay measures, upon which much of the research drawing enthusiasm for the delay of gratification construct is based.

Internal coherence of delay of gratification measures. Internal coherence and consistency in individuals' psychological measures are important for ensuring correspondence between empirical inquiry and theory. For instance, should delay of gratification performance reflect a unitary latent ability (i.e., self-control), then performance across repeated choices should exhibit a reasonable degree of consistency across items: delaying on one item should be positively related to delaying on another. Technically referred to as internal reliability, measuring this form of consistency is disallowed by the design of many delay of gratification tasks – including the infamous marshmallow test – by virtue of the fact that many include only one item. As administered across studies in the literature, choice delay tasks include one (e.g., Thompson et al., 1997) to nine (e.g., Prencipe & Zelazo, 2005) items. Internal reliabilities of choice delay assessments, when reported, are generally within adequate ranges (e.g., Schwarz et al., 1983), which is consistent with the notion of a single latent factor underlying choice delay performance.

However, simple internal reliability metrics are coarse and, on their own, do not represent sufficient evidence for uni-dimensionality of measure performance. Processes other than self-control are sometimes mentioned in the literature as influences on choice delay performance. For instance, children choose to delay gratification more often and more consistently with a shorter delay time. This may reflect greater certainty in promised rewards and/or greater ease for children conceptualizing a shorter time lapse (Schwarz et al., 1983). While certainly not conclusive evidence of multiple factors underlying choice delay performance, this finding highlights a need for a more sophisticated investigation into the psychometric properties of choice delay measures.

Construct validity of choice delay measures. Across childhood and many measurement paradigms, delay of gratification is positively related to stable temperamental and personality traits including effortful control (Davis, Bruce, & Gunnar, 2002), low activity level and low emotionality (Mittal, Russell, Britner, & Peake, 2013), and conscientiousness (Duckworth, Tsukayama, & Kirby, 2013). Moreover, delay of gratification is predicted by physiological activity (Sturge-Apple et al., 2016), is concurrently associated with physiological activity (i.e., heart rate reactivity, electrodermal response; Wilson, Lengua, Tininenko, Taylor, & Trancik, 2009), and predicts individual differences in functional imaging forty years later (Casey et al., 2011). Thus, delay of gratification has often been conceptualized as an aspect of temperament and is expected to relate to other self-regulatory measures. However, most prior research indicating that delay of gratification is tied to other self-regulatory or temperamental constructs is based on measures other than choice delay.

Choice delay performance relations with self-regulatory outcomes. Research focused specifically on choice delay performance has yielded less convincing evidence for it reflecting self-regulatory ability. In a meta-analysis including 236 studies, Duckworth and Kern (2011) evaluated the convergent and discriminant validity of delay tasks with other self-control related measures, including executive function (EF) tasks, and self-reported and informant-reported questionnaires. Delay tasks as a group were positively, but not significantly, correlated with EF tasks and were positively *and* significantly correlated with informant reports of self-control. Correlations for choice delay tasks with each other type of measure were generally weaker than for other types of delay tasks and were only significant for informant reports of self-control (Duckworth & Kern, 2011). Importantly for this paper, most of the studies included in the meta-analysis were based on relatively high-SES samples. Thus, despite putting forth a reasonable argument for delay tasks being at least partially related self-control, they do not address this issue specifically for low-SES children.

Weak relations between choice delay performance and other self-regulatory outcomes have several potential explanations as posited in the literature. Developmental period may be an important consideration, for example: whereas many studies found weak to null evidence for relations between choice delay and self-control or other related outcomes in young children (Bowersox, 2013; Hongwanishkul et al., 2005; Prencipe & Zelazo, 2005; Schwarz et al., 1983; Spann, 2014; Tynan, 2014), others have found significant relations between choice delay and risky behaviors in adolescents (Wulfert, Block, Santa Ana, Rodriguez, & Colsman, 2002). Others have suggested the direct measures for non-affective EF (e.g., inhibitory control) are relatively abstract and decontextualized, whereas the latter may be more context-bound (Bellagamba et al., 2015). Thus, it may be that contextual influences on choice delay performance complicate its interpretation and muddy its relations with other constructs. In conflict with this conjecture, Beck and colleagues (2011) offer that stable, “individual differences (e.g., temperament)” may underlie choice delay at an early age in order to explain the measure’s wide variance and lack of relations with other self-regulatory constructs. Finally, as noted by Honwanishkul and colleagues

(2005), who unexpectedly observed negative partial relations between choice delay performance and another "affective executive function" task, it may simply be that the conceptual basis for constructs involving both affective and self-regulatory processes need refinement.

***Predictive validity of choice delay tasks*.** Should delay of gratification indeed reflect children's self-control ability, or more general self-regulatory abilities, then children who delay more often should theoretically experience more success in settings expecting, rewarding, or requiring self-restraint. Formal schooling is one such setting, as children often face temptations and experience impulses which may interfere with their academic success in the long-term. Engaging in playful distractions from peers may be enticing in the short-run but, over time, may result in disciplinary action and poor rapport with one's teacher and/or get in the way of completing and benefiting from classwork. Indeed, self- and informant-reports of self-control are strongly related to outcomes such as grade point average (Duckworth, Tsukayama, & Kirby, 2013; Duckworth, 2011).

Choice delay performance in young children is a much less consistent predictor of school-related outcomes than sustained delay performance. Multiple dissertation studies have found null relations between choice delay performance and positive classroom behaviors (Tynan, 2014), parent reports of attention and aggression (Spann, 2014), and teacher reports of social competence, aggression, and opposition (Bowersox, 2013). In Bowersox's (2013) high-risk sample, choice delay performance was actually *negatively* related to measures of mathematics performance, although this relation was not robust enough to hold after multiple comparison adjustment. The latter point is interesting in light of the emerging theory outlined above regarding the potential that delaying gratification may not be an adaptive outcome in some populations.

Summary. In short, evidence for the validity of choice delay performance, both as a measure reflecting self-control processes and as a predictor of other outcomes, is weak; this is in contrast to a reasonable body of evidence supporting the same claim for sustained delay performance. And yet, choosing to delay gratification is a precondition for sustaining delayed gratification (Reynolds & Schiffbauer, 2005), and so proclivity to choose immediate over delayed gratification should theoretically impede delaying gratification, even in situations involving a waiting period. The under-representation of children of color from low-income families in delay research is problematic, given potentially different processes underlying delay in this population (e.g., Price-Williams & Ramirez, 1974; Strickland, 1972; Sturge-Apple et al., 2016). It is important, then, to reconsider whether the tendency toward choosing to delay gratification – independent of one's ability to sustain said choice(s) – is related to self-regulation and related outcomes in this population.

Present Study

The present study examines the psychometric properties of a choice delay measure in a sample of mostly Black (some Latino/a) kindergarteners from low-income families living in an urban school district. Specifically, I address three research questions:

Research Question 1: Is performance on the choice delay measure internally consistent and coherent in the present sample?

Presuming a single, latent construct underlies choice delay performance, children's performance across all six items should exhibit reasonable consistency (i.e., reliability). Poor consistency may indicate different factors underlying performance across different reward types, for example, and/or may warrant consideration of new scoring schemes.

Thus, the present study delves into children's choice delay performance using item factor analysis – which is analogous to item response theory – in order to model not just consistency in performance across items, but also item-specific parameters, such as the difficulty and discriminatory value of individual items. Further, factor analysis allows a test of whether certain items reflect different latent factors, even if performance across items is still highly correlated, which could be consistent with different processes explaining performance across items. Calculation of difficulties also allows a test of coherence in performance: larger ratios between smaller-sooner and larger-later rewards are generally expected to be correlated with higher likelihood to delay gratification, as a higher-ratio means a greater pay-off to the cost of waiting, which in this study is constant across items. Similarly, children's individual preferences for different types of rewards (in this study, stickers vs. candy) may drive their valuation of said rewards, such that more immediately alluring rewards may make delaying gratification more difficult.

The utilization of multiple-item measures for choice delay enables examination of reliability – or, said differently, verification of coherence – within children's delay of gratification performance. However, furnishing "smaller sooner" rewards immediately on early items could theoretically affect performance on later items. This would represent a violation of the local independence assumption required in item response theory, which could be problematic for interpretation of difficulty and discrimination parameters. More explicitly, if a child chooses to receive a piece of candy immediately on an early item, he or she may find it easier to delay on later candy items, as the allure of candy may be reduced while or immediately after consuming the piece already received. It also has the potential to muddy the relation between, for example, reward ratios and item difficulty, as item order may also influence the latter. Therefore, the present study will consider item ordering as a potentially influential factor for both item difficulty and discrimination indices. Because item-ordering was fixed across all choice delay assessments, it is not possible with these data to determine with confidence the extent to which each item characteristic (e.g., type, ratio, order) influenced item difficulty. Instead, rough correspondence between difficulty and item characteristics may be interpreted as an indication of coherence in children's performance.

Even if performance across choice delay items significantly and monotonically relates to a single factor representing a single process (e.g., self-control), it is possible that not all items equally reflect that factor/process. Item factor analysis will provide estimates for this discriminatory power of each item in addition to difficulty, which will inform whether certain items are more strongly related to the common latent factor(s) underlying performance than others. Variability in discrimination indices across items would be consistent with the notion of alternative processes influencing children's performance, with those processes more influential on some items than others.

Potential differences in the way choice delay items operate (i.e., difficulty of items, differences in fit of measurement models) across groups were tested in order to confirm that operation of the measure was consistent across said groups. Girls and boys were compared given some evidence of differences in delay proclivity (Spann, 2014) and in mechanisms underlying choice delay performance by gender (Toner et al., 1977). Children assessed during the summer versus during the school year were compared since the assessment settings differed somewhat based on timing, as did recruitment method. Children who scored at floor on the EF measure versus those who did not were also compared as a potential means to unpack the unexpected, negative correlation observed between choice delay and EF in the present sample (Duran et al., in press). Finally, as part of the sensitivity analyses focused on detecting interference in results by program participation (the focus of the broader study from which these data were derived), treatment and control groups were also compared.

Research Question 2: Is choice delay performance related to other, concurrently measured and theoretically-aligned constructs?

Being based on secondary data from a large-scale evaluation study, the present study will consider child-direct measures and teacher-reports of self-regulatory constructs and school-related outcomes. Theoretically-aligned constructs measured concurrently with choice delay included teacher reports of self-management and responsible decision-making and a direct measure of EF. Based on extant literature, weak, but positive correlations, should be expected between choice delay performance and teacher reports of self-management and responsible decision-making (Duckworth & Kern, 2011). On the other hand, choosing immediate gratification may be adaptive in some contexts (Ellis et al., 2017; Lee & Carlson, 2015) and this would be consistent with negative correlations between choice delay and EF observed by Duran and colleagues (in press) in the present sample. So, in this sample it may be the case that children who choose to delay gratification exhibit less self-management and responsible decision-making, as they tend to have lower cognitive self-regulatory ability.

As observed in Duran and colleagues (in press), I expected the negative correlations between choice delay and concurrently measured EF here, but expected partial correlations to be weaker due to the inclusion of language ability as a control (which was not included in Duran et al., in press). It may be the case, for instance, that low language ability precluded some children from understanding one or both of these measures, which may explain the unexpected, negative correlation previously observed.

Research Question 3: Does choice delay performance predict teacher-reports of children's classroom behaviors and direct measures of academic achievement?

Available measures of relevant classroom behaviors in the present study's data set included teacher reports of difficulties transitioning to kindergarten, problem behaviors, and positive classroom behaviors. Difficulties transitioning to kindergarten was measured at the beginning (i.e., in the fall) of the school year; all other outcomes were measured at the end (i.e., in the spring) of the school year – hence, their use in these predictive analyses as opposed to as concurrent measures of self-regulatory constructs under research question #2. Classical theory would predict that choice delay would positively predict positive classroom behaviors (self-control and self-regulation); prior empirical evidence is consistent with this notion but suggest only weak relations should be expected (Duckworth & Kern, 2011). Choice delay should therefore also negatively predict difficulties transitioning to kindergarten and first grade and problem behaviors (externalizing behaviors and hyperactivity).

However, as previously mentioned, emerging theory (e.g., Ellis et al., 2017) and the negative correlation between EF and choice delay in the present sample (Duran et al., in press) left open a reasonable possibility that children who tended to delay gratification also had more difficulty transitioning to new classroom contexts. Should this be the case, it would not necessarily *have* to be the case that the same children exhibiting high delay and also having difficulty transitioning to the school year must therefore also exhibit less positive and more problem behaviors at the end of the year. First, high tendency to delay gratification may indicate socialization in values well-aligned with the school system. But, as high-delaying children in this sample tend to have low EF, they may struggle to enact these values in the face of novel cognitive and behavioral demands at the beginning of the school year. By the end of the school year, once they have acclimated to said demands, their self-control ability (or socialization) may be more evident to their teachers. Second, the correspondence between low tendency to choose to delay gratification and high EF in this sample may be an artifact of novelty of the testing scenario for children – a possibility discussed by Duran and colleagues (in press). Therefore, delay of gratification tendencies may not have a strong correspondence to behaviors over a longer timeframe, as reflected in the end-of-year teacher reports.

Academic achievement measures were tested at the beginning of kindergarten, beginning of first grade, and beginning of second grade. Theoretically, delay of gratification is not necessarily related to children's academic skills and knowledge prior to entry to formal schooling, and so choice delay performance is not expected to significantly relate to concurrently measure academic achievement. However, choice delay may relate to later achievement and improvement over time, consistent with the notion that delay of gratification behaviors reflect abilities that support adaptive learning behaviors in school contexts.

Method

The present study is based on secondary data analysis utilizing data from an evaluation study of an after-school, social and emotional learning (SEL) program held in an urban elementary school district in a southeastern state of the U.S. As part of the design of the evaluation study, some children were randomly selected to have the opportunity to attend the aforementioned SEL program, and others were not; impact analyses are forthcoming. As the program's goal was to foster many social and emotional competencies related to the classroom behaviors and academic measures considered in the present study, sensitivity analyses were conducted to ensure there was no interference of program participation with results.

Data/Participants

Participants in the present study were 371 children (47% male) recruited from four elementary schools in an urban school district hosting the after-school, SEL program. Three consecutive cohorts of children were recruited and included in the study, with the first time point for data collection for each cohort occurring prior to or at the beginning of kindergarten (Time 1; M_{age} = 5.4 years). Data collection for measures of interest in the present study continued through the beginning of second grade.

Demographic characteristics of study children (see Table 1) were generally representative of the high-poverty population served by the school district. Among children for whom demographic information was available at baseline, 96% were receiving free or reduced-price lunch, 81% of children's families were also receiving some other form of public assistance, and 54% were living in a household with only one adult. Fifty three percent of children were female; 88% were Black and 4.5% were Hispanic/Latinx (remainder Multi-race or unspecified).

Procedure and Measures

During the summer of each year of the study, all study children had the opportunity to attend a free, week-long "camp," during which trained research assistants administered multiple direct assessments to each child in three, 15- to 20-minute batteries. Children who did not attend the summer camp were tested in a quiet space in their school at the beginning of the school year. At the same time points as the child assessments, parents (89% mothers) were surveyed for demographics, among other information. Surveys were administered to parents during the week of summer camp if they were available, but otherwise took place over the phone.

Choice delay of gratification. In the choice delay of gratification task, researchers offered children a series of choices between a small, immediate reward (e.g., one sticker) and a larger reward (e.g., four stickers) which would be accessible (in their backpacks) at the end of the day. Two types of rewards were used (i.e., stickers and candy) and three smaller-sooner to larger-later ratios were used (i.e., one to two, one to four, and one to six) for a total of six items. These six items were administered in the same order for all children at all time points:

1. One sticker now or six stickers later
2. One piece of candy now or two pieces of candy later
3. One piece of candy now or four pieces of candy later
4. One sticker now or two stickers later
5. One sticker now or four stickers later
6. One piece of candy now or six pieces of candy later

In main analyses, children’s choice delay performance was modeled as a single latent factor with the six items as indicators. In sensitivity analyses, alternative scoring schemes included (1) a composite sum (count) of the number of items on which children chose to delay (range: 0 to 6); (2) performance on the first item only (delay = 1; now = 0); and (3) dichotomized performance dichotomized across six items (“high delay” = delaying on more than three items). Dichotomized performance has been used in some prior studies (e.g., Göllner, Ballhausen, Kliegel, & Forstmeier, 2017; Lee & Carlson, 2015) and is potentially prudent in the present sample, given that the distribution of choice delay performance was bimodal (see Figure 1). Internal reliability (Cronbach’s alpha) across the six items in the present sample was .83.

Concurrently measured self-regulation. Measures of executive function (EF), responsible decision-making, and self-management were used as self-regulatory measures in assessing construct validity. EF was measured using the Head-Toes-Knees-Shoulders (HTKS) task. This three-part measure requires children to do the “opposite” of verbal commands. In part one, children must touch their head when told to touch their toes, and vice versa. In part two, children must utilize the rule from part one, in addition to touching their knees when told to touch their shoulders, and vice versa. In part three, rules switch (e.g., children must touch their knees when told to touch their head). Performance on this measure simultaneously taps multiple components of EF and has strong validity as a self-regulatory construct and predictor of school outcomes (Cameron et al., 2012; Cameron Ponitz, McClelland, Matthews, & Morrison, 2009). Internal reliability (Cronbach's alpha) in the present sample was .90.

Teacher reports of SEL competence were collected at the beginning of kindergarten with the Devereux Students Strengths Assessment (DESSA), a five-point rating scale which has been well-validated for school-aged children through eighth grade, including in the present sample (Dormoral, Cottone, & Kim, 2017). The two of five subscales used as concurrent measures of self-regulation were self-management and responsible decision-making. The self-management subscale measures children's successful management of emotions and behaviors across many situations, such as when waiting one's turn or adjusting to changes in plans. The responsible decision-making subscale measures children's problem solving and constructive decision-making in social situations, as well as their ability to learn from prior personal experiences and/or those of others, while taking into account factors such as ethics and norms in guiding behavior. A recent study provided evidence that the DESSA, administered in the present sample, measures five distinct, but highly related, competencies, with strong relations to other directly-measured and teacher-reported school outcomes (Dormoral et al., 2017).

Classroom Behaviors. Teacher reports of difficulties transitioning to kindergarten were collected at the beginning of kindergarten and the beginning of first grade with the Survey of Early School Adjustment Difficulty (Rimm-Kaufman, 2005). This 11-item measure, designed and validated for use in kindergarten and first grade classrooms, focuses on children's ability to adapt to the classroom, with a majority of items corresponding to either social skills (e.g., "This child has shown difficulty getting along with other children") or self-regulatory ability (e.g., "This child has shown difficulty taking turns or waiting until his/her turn to speak"). Internal reliability in the present sample was .93 in both kindergarten and first grade.

Teacher reports of classroom behaviors were collected at the end of both kindergarten and first grade with the Social Skills Improvement System (SSIS; Gresham & Elliott, 2008) and the Child Behavior Rating Scale (CBRS) Self-Regulation subscale (Bronson, Goodson, Layzer, & Love, 1990). Both scales are four-point rating scales measuring the frequency of specific behaviors (1 = "never" to 4 = "almost always"); are well validated for use in classroom settings (Bronson et al., 1990; Crosby, 2011; McClelland & Morrison, 2003). The SSIS subscales to be used in assessing predictive validity include externalizing and hyperactivity (problem behaviors) and self-control (a positive classroom behavior). Internal reliability on each of these subscales at each time point ranged from .80 to .95. A negative classroom behaviors composite was created by averaging scores on the SSIS externalizing and hyperactivity subscales; a positive classroom behaviors composite was created by averaging scores on the SSIS self-control and CBRS self-regulation subscales.

Academic Achievement. Academic achievement was assessed at the beginning of kindergarten, beginning of first grade, and beginning of second grade with the Woodcock Johnson Tests of Achievement (WJ-Ach III). Analyses utilized as a composite an average of scores on the Applied Problems (mathematics), Letter-Word Identification (literacy), and Academic Knowledge (social studies and science) subtests.

Language. Language ability will be used as a covariate in predictive validity analyses and was assessed using the Differential Abilities Scale (DAS) Naming Vocabulary subscale. Language is covaried out in many other studies focused on delay of gratification (e.g., Sturge-

Apple et al., 2017). The DAS has exhibited good validity as a tool for identifying preschoolers as at-risk or in need of supports (McIntosh, 1999).

Data Preparation – Missing Data Handling.

All missing covariates and outcomes for each analytic model were imputed using multiple imputations with chained equations, carried out in the `mice` package (Van Buuren & Groothuis-Oudshoorn; 2011) in R version 3.2.5. Two separate imputations were carried out: one for data required to answer the first two research questions, and one for the third research question. Each set of imputations comprised 20 imputed data sets and a maximum of 20-40 iterations for each imputation set, which are generally considered adequate (Enders; 2010; Van Buuren and Groothuis-Oudshoorn, 2011). In all sets of imputations, item-level choice delay performance was imputed using logistic regression, and total choice delay performance was then derived from the imputed item-level performance. Teacher reports and academic achievement were imputed using predictive mean matching at the subscale level (e.g., decision-making, applied problems); composites were derived from subscales thereafter.

In building the imputation model for each set of imputations, all analytic variables were required to be used as predictors for all other analytic variables. Additionally, the `quickpred` function of the `mice` package was used to identify sets of useful predictor variables among other demographic information and child direct assessments. Finally, analytic variables from all sets of imputations were included for use as potential auxiliary variables in all other imputation sets in order to ensure the program had consistent information available across all sets of imputations.

Data Analysis

Analysis was carried out in Mplus v. 7.0 (Muthén & Muthén, 2012). Models including dichotomous choice delay items or other scoring schemes (e.g., high vs. low performance) as outcomes used robust mean- and variance-adjusted weighted least squares estimation (Brown, 2006). All other models based on alternative scoring schemes used maximum likelihood estimation. In models predicting school-related outcomes, clustered standard errors based on teacher identification numbers were used to account for dependence in children's scores as a result of shared classroom experiences and a shared reporter (i.e., their teacher).

Research question 1 – internal consistency and coherence of choice delay performance. Internal consistency (reliability) and coherence of children's choice delay performance at the item-level (RQ #1) were assessed within an item factor analysis (IFA) approach (Wirth & Edwards, 2007). A categorical confirmatory factor analysis (CCFA) was run testing a single-factor model, with all six items loading onto the same latent factor. Loadings for all six items were freely estimated and the variance of the latent factor was set to one. Model fit was assessed based on the comparative fit index (CFI) and Tucker-Lewis Index (TLI), as described by Cheung and Rensvold (2002). Then, a Rasch model scoring scheme (with equal loadings across items) was tested for adequate fit and compared to the unconstrained CCFA to ascertain whether data were consistent with items equally reflecting a single, underlying factor.

Multi-group analyses were used to determine whether there were differences in the way the choice delay items operated between potentially relevant subgroups: girls versus boys; children assessed during summer versus those assessed during the school year (corresponds roughly with recruitment method); and children who scored at- versus above-floor on HTKS.

Assumptions check: dimensionality and local independence. Furnishing "smaller sooner" rewards immediately on early items could theoretically affect performance on later items. More explicitly, if a child chooses to receive a piece of candy immediately on an early item, he or she may find it easier to delay on later candy items, a violation of the local independence assumption required in item response theory. Local item dependence (LID) can affect reliability estimates and has the potential to muddy the interpretation of item parameters (see Zenisky, Hambleton, & Sireci, 2001). Another potential form of LID which could be of concern here is closer relations between certain items than others based on reward type, as children's preferences regarding delaying or not may depend on reward type. Therefore, I conducted formal and informal tests of local item dependence (LID). These included pairwise correlations between the residuals of each item after accounting for the latent factor (also known as Q_3 scores; Yen, 1984); CFAs testing alternative factor structures based on item characteristics (reward type and ratio); and examination of response pattern frequencies to determine whether certain patterns were more common than others or than what would be expected based on chance. Of particular interest were patterns in which children chose immediate gratification on early items and chose to delay on later items, given the potential for immediate furnishing of early rewards to affect salience of later rewards. These analyses complement the sensitive analyses, described later, assessing the relative construct validity of different operationalizations of performance (e.g., performance on the first item only).

Research question 2 – concurrent relations with other self-regulatory outcomes. Zero-order and partial correlations were used to determine whether the choice delay performance was related to measures of other, theoretically-aligned, concurrently-measured constructs (RQ #2). These included teacher reports of self-management and decision making (DESSA) and a direct measure of executive function (i.e., HTKS). Choice delay performance was modeled using the latent factor (as in RQ #1). Partial correlations controlled for age, gender, language ability, and assessment timing.

Research question 3 – predictive validity. Regression models were used to understand the extent to which children's choice delay performance (again, modeled as a latent factor) predicts classroom behaviors and academic achievement (RQ #3). Separate models were run for difficulties transitioning to kindergarten, positive classroom behaviors and problem classroom behaviors at the end of kindergarten, academic achievement, and improvements in academic achievement. Separate models were run for positive classroom behaviors (i.e., self-control and self-regulation) and problem classroom behaviors (i.e., externalizing and hyperactivity) because reports of desirable and undesirable outcomes do not always correspond. For academic achievement, the model of central interest was one predicting achievement at the beginning of first grade controlling for baseline achievement (i.e., improvement in achievement). However, in order to fully understand relations between choice delay and academic achievement, two other models were also examined – one predicting baseline achievement, and another predicting achievement at the beginning of first grade (not controlling for prior achievement).

In order to understand whether the predictive validity of the choice delay measure holds when considering outcomes another year into the future, the analyses just described were repeated on the same outcomes when collected one year later (e.g., transition to first grade instead of transition to kindergarten).

Covariates in regression models for all outcomes included children's age at choice delay assessment, choice delay assessment timing (summer vs. school year, plus timing in days within each assessment context), gender, mother's education, mother's age, cohort, school, and assigned treatment condition. Mother's education and age were included given evidence that adult tendencies to delay gratification and discount future rewards vary with education and age

(Göllner et al., 2017) and, therefore, it is possible that these variables are related to socialization of children's delay preferences. Both are also well-known correlates of school-related outcomes.

Further, models were compared including and excluding baseline HTKS performance as an independent variable. While EF and delay of gratification are theoretically related as self-regulatory abilities, empirical research suggests weak-to-null relations between performance on measures of each construct (Duckworth & Kern, 2011). However, as discovered in Duran, Cottone, Ruzek, Mashburn, and Grissmer (in press; i.e., Paper 2), choice delay and executive function (i.e., HTKS) are negatively related at baseline in the present sample. Thus, in this sample in particular, it may important to understand the extent to which EF is confounded (or not) in the predictive relations between choice delay performance and these outcomes of interest.

Sensitivity Analyses

In exploratory, sensitivity analyses, different scoring schemes for choice delay were used to explore the second two research questions in order to understand whether a particular scheme showed strongest relations with other outcomes. These were: a continuous mean (as in most research focused on this measure; Prencipe & Zelazo, 2005), a dichotomized indicator of "high" versus "low delay," and performance on the first item only. The exploration of first item performance as a scoring scheme was motivated by the use of single-item choice delay measures in other studies (e.g., Thompson et al., 1997). A single item score also follows the rationale that performance on later items may be dependent on performance on early items, and therefore the former may not be as meaningful as the latter.

Sensitivity analyses were also conducted in order to ensure that participation in the after-school program, the evaluation of which is the focus of the broader study from which the present data were derived, did not interfere with or drive my results. In these sensitivity analyses, the central analyses described under the analytic plan were run in a multi-group model, with assigned treatment condition as the grouping variable. Each multi-group model was run once as an unconstrained model, and once with parameters of interest to be constrained between treatment and control groups. As suggested by Cheung & Rensvold (2002), a change greater than .01 in the CFI was interpreted to indicate that parameters of interest differed between treatment and control groups. In practical terms, this would also have indicated that the opportunity to participate in the after-school program drove differences in relations among constructs and would therefore call into question the external validity of the present results.

Results

Univariate descriptive statistics for demographic and analytic variables are shown in Table 1. Performance across choice delay items was somewhat uniform across items, with at least 47% (for Item 2) and at most 57% (for Item 3) of children choosing to delay. The average choice delay score was 3.14 out of 6, which corresponds to children delaying on 52% of items, on average. Very few children represent this “average” score, however, as choice delay performance across items was bimodally distributed (see Figure 1). The two most common scores were six and zero items, earned by 20% and 19% of children, respectively; the least common scores were two and three items, earned by 9% and 10% of children, respectively. Forty-nine percent of children delayed on more than three items (better than chance).

Bivariate correlations between demographic and control variables and choice delay performance (see Table 2) were largely non-significant. The strongest correlate of performance was assessment timing, with summer assessment being associated with higher performance than assessment during the school year ($r = .17$, $p < .001$ for composite choice delay performance). Naming vocabulary was also correlated with multiple measures of choice delay performance, with greater vocabulary being associated with delaying on fewer items (*r* = -.11, *p* = .05 for composite choice delay performance).

Research Question 1: Is the choice delay measure internally consistent (reliable) and coherent in the present sample?

As shown in Table 3, an unconstrained, single-factor, CCFA model exhibited adequate fit across all indices and imputed data sets, with CFIs of at least .99 and TLIs of at least .98.

Rasch model versus an unconstrained CCFA. As also shown in Table 3, the Rasch-model (with equal loadings across items) exhibited poor fit on all indices, with CFIs and TLIs no greater than .94 across the 20 imputed data sets. As these represent a drop in CFI and TLI greater than .01 (Cheung & Rensvold, 2002), I conclude that the Rasch model fit the data significantly worse than the unconstrained CCFA. Therefore, hereon forward, all analyses including choice delay performance modeled as a latent factor utilize and refer to an unconstrained CCFA.

Parameter estimates for single-factor CCFA. Parameter estimates for the single-factor CCFA are shown in Table 4. Tetrachoric correlations were estimated because the response variables were binary; these revealed moderate to strong relations among items, with the lowest correlation between Item 1 and Item 6 (*r* = .47) and the highest correlation between Item 3 and Item 6 (*r* = .78). Further, standardized factor loadings were all significant, ranging from .68 for Item 1 ($p < .001$) to .87 for Item 3 ($p < .001$), suggesting all items strongly related to overall performance. Taken together with the adequate fit of the model (as previously described), these observations are consistent with the notion of a single factor underlying performance on the choice delay measure.

The range of thresholds for items was somewhat narrow, from -.17 ($p < .05$) for Item 3 to .09 (*p* = .16) for Item 2, with four out of six thresholds not significantly different from zero, indicating that these items were of "average" difficulty. In other words, a child of average choice delay "ability" had a roughly 50% chance of delaying on each item with a threshold of zero. The threshold of -.17 ($p < .05$) for Item 3 indicates that the item was slightly easier than average, such that 50% of children with .17 SD lower than average ability can be expected to delay on the item. Similarly, the marginally significant threshold of -.11 ($p = .10$) for Item 5 indicates that 50% of children with .11 SD lower than average ability can be expected to delay on the item. Given this narrow range of thresholds, planned comparisons of difficulties across items by characteristic (e.g., by small-to-large-reward-ratio) would not be meaningful.

Multi-group comparisons testing invariance. For each multi-group analysis, models constraining item loadings, measurement intercepts (thresholds), and residual variances were compared to incrementally test metric invariance, scalar invariance, and strict factorial invariance, respectfully. Strict factorial invariance implies that group differences in the residual variances of items are due to variances in the latent variable. Further comparing the fit of the strict factorial model to a model constraining latent means to be equal provides a valid test for observation of latent mean differences between groups. Briefly, these comparisons suggested strict factorial invariance, including equal measurement intercepts and residual variances, across all subgroups. Latent variable means were also equal across subgroups except those distinguished by assessment timing: children tested in the summer scored .39 SDs ($p < .01$) higher than those tested during the school year.

A note on missing data handling in multi-group analysis: Multigroup models comparing male and female children and those comparing treatment and control children utilized imputed data. Those comparing children who earned zero and non-zero scores on HTKS and those comparing children who were assessed during the summer versus during the school year were carried out using listwise deletion, as differing numbers of children in each subgroup across imputed datasets disallowed pooling imputed results.

Testing dimensionality and item independence. Other CCFA models were run to determine whether modeling performance in terms of two factors (one for each reward type) or three factors (one for each reward ratio) might fit the data better than a single factor model. Fit indices for the two-factor, "reward type model" were very similar to the single-factor model, with both CFIs and TLIs at least .994 and .988, respectively, across all imputed data sets. The two factors were highly correlated, however (r = .95, p < .001); this fact, combined with the observation of little degradation of fit from between the two models, is consistent with the single factor adequately capturing performance across items. Identification of the three-factor, "reward ratio model" required fixing one item's loading per factor equal to one in addition to constraining the variance of each latent factor to one. Leaving loadings unconstrained led to correlations greater than one between latent factors, indicating a high degree of collinearity in the model. Fit indices for the final model were excellent, with CFIs and TLIs of at least .997 and .993, respectively, across imputed data sets. Further, all item loadings onto each respective factor were significant. However, as with the "reward type model," the factors in the "reward ratio model" were highly correlated, with correlation coefficients between .77 (p < .001) and .82 (p < .001). In short, these CFAs provided inadequate evidence to support modeling choice delay performance as a construct with separable dimensions by reward type and/or reward ratio.

Correlations between the residuals of each item after accounting for the latent factor were were positive and significant between the residuals of item 1 and item 2 (r = .26, p < .05) and between the residuals of item 3 and item 6 (r = .31, p < .05). Correlations were negative and significant between item 1 and item 6 (r = -.38, p < .05). As 15 separate correlations were run, however, none of these correlations were significant following the Benjamini-Hochberg procedure for multiple comparison adjustment.

Table 5 shows the 20 most common response patterns, along with their observed frequency. The five most frequently observed response patterns represented over half (51.7%) of the sample, starting with delaying on all items or zero items, representing 71 and 67 children (20.1 and 18.9% of the sample), respectively. The three next most common response patterns were either delaying on *item one only* (16 children, 4.5% of sample), delaying on *all items except item one* (15 children, 4.2% of sample), or delaying on *all items except item two* (14 children, 4% of sample). These rates are 2.9, 2.7, and 2.5 times the rate of unconditional chance and about twice chance rates after conditioning on receiving a score of either one (n = 45) or five (n = 57).

Given the interest in whether immediate gratification on early items was more common than on later items, the response pattern in which children delayed on *all but the first two items* was also examined. This response pattern was the sixth most common, representing 12 children, or 3.4% of the entire observed sample and 26.1 % of the 46 children who delayed on exactly four items. Thus, among children with a score of four, children delayed on *all items except one and two* at a rate about four times greater than chance levels (6.7%). In contrast, among the other 34 children who received a score of four, only 1 to 5 children were represented by each of the remaining 13 observed response patterns corresponding to a score of four.

Research Question 2: Is choice delay of gratification performance related to performance on other measures of theoretically-aligned constructs?

Table 6 shows correlations between choice delay performance and other concurrently-measured self-regulatory outcomes. When modeled as a latent factor, choice delay performance was significantly and negatively related to each of these outcomes. Zero-order correlations utilizing the latent factor model were $r = -.21$ for choice delay performance with EF ($p < .001$), $r = -.16$ with Responsible Decision-Making ($p < .05$), and $r = -.14$ with Self-Management ($p < .05$). After partialling out age, gender, language ability, and assessment timing, the direction of the correlations was the same, with magnitudes changing only somewhat: $r = -.19$ with HTKS ($p < .01$), $r = -.18$ with Responsible Decision-Making ($p < .01$), and $r = -.16$ with Self-Management ($p < .01$).

Research Question 3: Does choice delay performance predict teacher-reports of children's classroom behaviors and direct measures of academic achievement?

Table 7 displays variance explained in outcomes, along with regression coefficients for choice delay performance (modeled as a latent factor) and HTKS performance in models predicting teacher reports of classroom behaviors. Table 8 displays corresponding information for models predicting direct measures of academic achievement, along with coefficients for Kindergarten academic achievement in models predicting improvement in achievement. Covariates included in all models were: age, gender, assessment timing, naming vocabulary, mother's education, whether the child's mother was less than 21 years of age when he or she was born, school, cohort, and treatment condition.

Across classroom behavior outcomes in kindergarten (top half of Table 7), HTKS and choice delay explained similar amounts of variance when entered separately, and each added about one to three percentage points of additional variance explained above covariates and the other predictor when entered simultaneously. For example, about 16% and 15% of variance in positive classroom behaviors reported at the end of kindergarten was explained by the HTKS only and choice delay only models, respectively. Twenty percent of variance in this outcome was explained by the model including both self-regulatory predictors. Regardless of whether HTKS was included in the model, choice delay performance positively predicted difficulties transitioning to kindergarten ($\beta_{CD\ Only}$ = .16, p < .01; $\beta_{CD+HTKS}$ = .13, p < .05), positively predicted problem behaviors at the end of kindergarten ($\beta_{CD\ Only}$ = .13, p < .05; $\beta_{CD+HTKS}$ = .11, p < .05), and negatively predictive positive behaviors at the end of kindergarten ($\beta_{CD\ Only}$ = -.18, p < .01; $\beta_{CD+HTKS}$ = -.15, p < .05).

The pattern of results was largely similar when predicting first grade classroom behaviors (bottom half of Table 7), although the strength of choice delay performance as a predictor was weaker. Choice delay only models explained between 15 and 16% of variance in outcomes, less than the 17 to 23% explained in HTKS models. Further, the models including both predictors explained only 1-2% more variance than the HTKS models. While the direction of the choice delay coefficients was the same when predicting first grade classroom behaviors as when predicting kindergarten classroom behaviors, the magnitude and confidence level of these coefficients was reduced. Choice delay positively, but not significantly, predicted difficulties transitioning to first grade ($\beta_{CD\ Only}$ = .08, p = .23; $\beta_{CD+HTKS}$ = .07, p = .34); positively predicted problem behaviors at the end of first grade ($\beta_{CD\ Only}$ = .13, p < .05), but only marginally so when including HTKS ($\beta_{CD+HTKS}$ = .10, p = .06); and negatively predicted positive behaviors at the end of first grade ($\beta_{CD\ Only}$ = -.18, p < .001; $\beta_{CD+HTKS}$ = -.14, p < .05).

Across measurement time points for academic achievement, choice delay negatively predicted achievement, with regression coefficients of β = -.10, -.13, and -.11 for beginning of kindergarten, beginning of first grade, and beginning of second grade, respectively (all p's < .05). Choice delay performance did not significantly predict achievement at any time point after controlling for HTKS. Choice delay performance also did not predict either first grade or second grade achievement after controlling for kindergarten achievement (see right half of Table 8).

Sensitivity Analyses – Comparing Scoring Schemes.

Comparing correlations with concurrent measures of self-regulation across choice delay scoring schemes. Correlations between choice delay and concurrently measured self-regulatory constructs were negative and significant across all scoring schemes (Table 6). Dichotomized performance ("High delay" = delaying on more than three items) was most strongly, and still negatively, related to other constructs. "High delay" was most strongly related to Responsible Decision-Making, at r = -.26 and r = -.27 for zero-order and partial correlations, respectively (both p < .01). Correlations based on other scoring schemes (composite sum and first-item-performance-only) were also in negative and significant, ranging from r = -.18 to -.14 (p < .05).

Predicting school-related outcomes with alternative scoring schemes. Table 9 compares the amount of variance explained in models predicting each school-related outcome, each utilizing a different choice delay scoring scheme, as well as the coefficient for choice delay performance for each model. The first column of parameters corresponds to a model including covariates and HTKS (no choice delay performance); the second through fifth columns each correspond to a model including covariates and HTKS in addition to one of four alternative scoring schemes for choice delay performance. These scoring schemes are: choice delay performance as a latent factor (as previously described); choice delay performance as a composite sum across all six items, performance on the first item only; and a dichotomized score, with "high" delay defined as delaying on more than three items and "low" delay as three or fewer items. For each outcome, in order to highlight models with most predictive power, the parameters associated with model explaining the most variance in the outcome are bolded only when the choice delay performance coefficient was at least marginally significant.

Predicting classroom behaviors with alternative scoring schemes. When predicting classroom behaviors in kindergarten, similar amounts of variance were explained across scoring schemes, within a range of up to .8 percentage points for each outcome. The latent factor model explained the most variance for each outcome: 1.3 to 2.3 percentage points beyond covariates and HTKS. As with models utilizing a latent factor model for choice delay, across all scoring schemes, higher choice delay performance at least marginally predicted more difficulties transitioning to kindergarten (β = .06 to .26, $p < .10$), at least marginally predicted more problem behaviors at the end of kindergarten for all schemes except first-item-only performance (β = .05 to .20, $p < .10$), and significantly predicted less positive behaviors at the end of kindergarten (β = -.07 to -.29, $p < .05$).

The pattern of results when predicting first grade classroom behaviors was similar to those predicting the same outcomes in kindergarten. However, for both difficulties transitioning to first grade and problem behaviors at the end of first grade, the only scoring scheme which produced significant choice delay regression coefficients was that utilizing performance on the first item only. Compared to the covariates and HTKS models, the first-item-only model explained 5.4 and 4.8 percentage points more variance in difficulties transitioning and in problem classroom behaviors, respectively. Those who delayed on the first item had 0.49 SDs' worth more difficulties transitioning to first grade ($p < .001$) and exhibited .46 SDs' worth more problem behaviors ($p < .001$) compared to those who did not delay on the first item. In contrast, *all* scoring schemes produced significant, negative regression coefficients for choice delay performance when predicting positive classroom behaviors. These ranged from β = -0.46 (for the first-item-only model, $p < .05$) to β = -0.14 (for the latent factor model, $p < .01$). Across scoring schemes, the first-item-only model again explained the most variance in positive behaviors at the end of first grade: 5.1% beyond that of the covariates and HTKS model. Children who delayed on the first item exhibited .46 SDs' worth less positive behaviors at the end of first grade ($p <$ .001) than those who did not delay on the first item.

Predicting academic achievement with alternative scoring schemes. When predicting academic achievement, the choice delay models explained little variance beyond covariates and HTKS (.7 percentage points or less) and did not produce significant regression coefficients for choice delay performance.

Sensitivity Analyses – Testing for Interference of Experimental Study Design

Multigroup measurement models comparing treatment and control children exhibited adequate fit when loadings were constrained between groups, with an average CFI of .997 (range: .995-.999) and average TLI of .996 (range: .993-.999) across imputed data sets. Constraining measurement intercepts, residual variances, and latent variable means across groups did not degrade fit – actually, improved fit indices – and therefore yielded adequate fit (average CFI: .992, range: .990-.995; average TLI: .992, range: .990-.995). Thus, the data were consistent with strict factorial invariance and equal latent means between treatment and control children across all imputed data sets.

Multi-group tests for invariance in results for research questions 2 and 3 therefore compared the model in which the measurement properties of the choice delay measure were constrained to models in which respective parameters of interest were constrained between groups. Constraining parameters of interest never reduced the average CFI of the model (across the 20 imputed data sets for each model) by more than .009. Thus, these analyses do not produce evidence that program participation interfered with the relations observed among constructs.

Discussion

The present study investigated the psychometric properties of a choice delay of gratification assessment administered at the start of kindergarten (M_{age} = 5.X years) to young children of color from low-income families. Performance across items exhibited high internal reliability, and results of an item factor analysis were consistent with the notion of a single factor underlying children's performance across items, though not all items equally reflected said factor. Children with a high proclivity toward delaying gratification had lower EF and were rated lower by their teachers on behavioral measures than their peers in both kindergarten and first grade. A similar trend was observed with respect to academic achievement, with choice delay negatively predicting achievement but not after controlling for its negative relations with EF and/or controlling for concurrent achievement.

Choice Delay Performance was Internally Consistent and Coherent when Modeled by a Single Latent Factor

Factor analyses and bivariate correlations between item residuals (after accounting for choice delay performance as a latent factor) did not yield robust evidence for multi-dimensionality of choice delay performance. This observation is consistent with the notion that a single factor explains similarities in performance across items. However, the bivariate correlations yielded some evidence for inter-item dependence prior to multiple comparison adjustment. Further, less formal consideration of common response patterns revealed that performance on items 1 and 2 deviated from chance levels among children delaying on all but one or two items. These observations are consistent with theoretical concerns regarding how item dependencies may arise, and specifically with the notion that immediate gratification on early items might make delaying easier on later items. Taken together, these observations do not provide strong evidence against modeling choice delay performance with a single-factor model but support as prudent the consideration of scoring schemes other than continuous sums in analyses assessing construct validity.

Item parameters from the single-factor CCFA suggest most items differed in terms of the extent to which they reflected the latent factor but did not differ from each other in terms of difficulty. Superior fit of the unconstrained model over a Rasch model (in which discriminatory indices across items are constrained to be equal) further reinforced the former point. Differing discrimination indices across items suggests that use of traditional reliability indices, which give all items equal weight, may not be valid. That items were roughly equal in terms of difficulty is contrary to the notion of items' ratios between smaller-sooner and larger-later rewards as a source of variance in the extent to which children are able to delay. It may be that item order was an alternative influence for both discrimination and "difficulty": performance on the last four items was better explained by the latent factor than the first two items, and there was an insignificant trend such that each of the last four items were easier than the first two. Thus, difficulty estimates are consistent with the notion that children found delaying easier after "hedging their bets" by choosing immediate gratification on the first two items. This might also explain why relations with the single, underlying factor may have been weaker on the first two items, as strategies across items drove different choices for these items.

Choosing to Delay More Often Corresponded with Less Desirable Ratings on Other Outcomes

Children who chose to delay gratification on more items were rated lower, on average, by their teachers in self-management and responsible decision-making and scored lower on a direct measure of executive function than those who chose to delay on fewer items. They also had more difficulties transitioning to kindergarten and exhibited fewer positive and more negative classroom behaviors, as reported by both their kindergarten and first grade teachers. Choice delay performance was also negatively related to academic achievement at the beginning of kindergarten, first grade, and second grade, but not after accounting for its negative relations with concurrently-measured executive function or, when predicting first and second grade achievement, concurrent academic achievement. Patterns of results were similar across alternative scoring schemes for choice delay performance, though performance on the first item only exhibited the strongest relations with outcomes measured at later time points.

These observations are in contrast to the null or weak, but positive, relations between choice delay and other outcomes (Lindstrom & Shipman, 1972; Prencipe & Zelazo, 2005; Schwarz et al., 1983; Toner et al., 1977). Only one other choice delay validation study has focused on a population as economically disadvantaged as the present sample, though, and found negative relations between choice delay performance and measures of vocabulary and mathematics achievement (Bowersox, 2013). These relations were not highlighted by Bowersox (2013), however, because they were either not robust to controlling for covariates, or because choice delay did not explain significant variance above and beyond covariates. The present results and those of Bowersox (2013) dovetail well with Sturge-Apple and colleagues' (2016) observation of *negative* relations between physiological indicators of self-regulation in toddlerhood and sustained delay of gratification performance in preschool among low-SES children (Sturge-Apple et al., 2016). Taken together, the results of these three studies suggest that a proclivity toward choosing immediate gratification may be normative as an indicator of adaptive development among low-SES children.

Alternative processes underlying choice delay differing by SES and race or

racial/ethnic minority status. Processes other than self-control have been put forth in the literature as potential confounds in children's choice delay performance. These may obscure the relation between choice delay and other child outcomes as currently reported in the literature. Explaining the *negative* relations observed in the present study may further require a consideration of these alternative processes specifically from the perspective of children from low-SES families.

Task comprehension. Aside from self-control ability, choosing to delay gratification could be due to children's failure to fully appreciate the choice put before them, as a result of limitations in either language ability or impulse control (Addessi et al., 2014; Paglieri et al., 2015). Said differently, misunderstanding of the choice and/or inability to inhibit a "go for more" impulse could lead a child to choose a larger-later reward without fully appreciating that this option comes at a cost in the form of a wait/delay. Given well-established differences in language ability by SES in early childhood (Hackman & Farah, 2009; Hoff & Tian, 2005; Sarsour et al., 2011), language ability may be more of a barrier to understanding the choice delay task among low-SES children in the present sample than that of prior studies. Further, it could be that higher-SES children are more familiar with scenarios in which adults are giving explicit choices between material rewards by virtue of the fact that higher-SES parents may have the luxury of offering children such choices. So, again, while comprehension may muddy relations between choice delay and other outcomes in middle-SES samples, it may represent a more systematic confound in extremely low-SES samples, as in the present study.

If comprehension of the task *was* a systematic challenge for participants in the present study, children may have accrued points on choice delay tasks as a result of language limitations or even difficulties with impulse control. This may at least partially explain the negative relations observed here between choice delay and other outcomes. Mistakenly choosing a larger-later reward on early items could also logically lead to the child delaying on later items after learning that they will not, in fact, receive that reward immediately. This could also be one explanation the high occurrence of response patterns in the present sample characterized by delaying on all but the first and/or second items.

Reward expectancy. Second, choosing *immediate* gratification on a choice delay task could be due to low expectancy for the larger-later rewards presented by experimenters (Freire, Gormana, & Wessman, 1980; Kidd et al., 2013; Lee & Carlson, 2015; Mahrer, 1956; Michaelson et al., 2013; Michaelson & Munakata, 2016). In the context of choice delay tasks, choosing larger-later rewards involves parting ways with the smaller-sooner reward and the researcher promising them. This scenario may be likely to registered as risky among lower-SES children, given differences in exposure to stress and instability by SES (Evans, 2004; Moore et al., 2000) which may affect their confidence in future events (Ross & Hill, 2002). Importantly, all research assistants administering the delay of gratification tasks in the present study were White and, at the time of the first time point of data collection, were fairly unfamiliar to children. This feature of the study may have intensified the role of expectancy in children's performance (Price-Williams & Ramirez, 1974; Strickland, 1972).

As greater cognitive flexibility is related to children downwardly adjusting their delay preferences in response to risk of losing later rewards (Lee & Carlson, 2015), proclivity toward immediate gratification in the context of low expectancy may represent strong, rather than weak, self-control ability. Thus, low expectancy for later rewards could explain the negative relations observed here between choice delay and other self-regulatory constructs here. It may also explain why a disproportionate number of children in the present study chose immediate gratification on either item one or two *only* – doing so could have been a strategy in which children with greater flexibility in their preferences “hedged their bets” on later items.

Reconciling choice and sustained delay research. As sustaining delayed gratification is predicated on first choosing to delay, the alternative processes underlying choice delay should therefore also be confounded in sustaining delay. A child will not resist impulses to consume a smaller-sooner reward if he or she has not first deemed that doing so will be beneficial. The null relations between choice delay and other outcomes in middle- to upper-SES samples observed in the literature are interesting, then, given the robustly positive relations observed in such samples for sustained delay tasks.

Also interesting are the relatively weak findings of a recent replication of the original “marshmallow test studies” focused on a predominantly low-SES sample (Watts, Duncan, & Quan, 2018). The relations between sustained delay at age four and academic achievement in adolescence were positive, but much smaller in Watts and colleagues (2018) than those found in original studies (Shoda et al., 1990). Further, these relations were highly sensitive to the inclusion of child characteristics, weakening the argument that delaying gratification plays an operational role in supporting children’s success in life (Watts et al., 2018). Interestingly, though, the majority of variance explained in outcomes by time children waited in Watts and colleagues (2018) was attributable to waiting at least 20 seconds. This may suggest that, for these lower SES children, their initial judgment regarding whether or not they *should* delay was just as indicative of children’s later academic achievement in adolescence as *sustaining* their choice.

It may be that contextual differences between how choice and sustained delay are typically *measured* – aside from those that map onto the theoretical differences in the *constructs* – may explain the conflicting directions of results between those of Watts and colleagues (2018) versus those observed here and by Bowersox (2013). For example, it is unlikely that lack of comprehension on a sustained delay task could lead to greater likelihood of waiting for the larger-later reward. Further, as children do not need to part ways with smaller-sooner rewards in sustained delay tasks, delaying on a sustained delay task may not seem as inherently risky to children who are attuned to such risks. Differences in the racial composition between the present study and Bowersox (2018) and that of Watts and colleagues (2018) which could further explain differences in results across studies, given differing processes underlying delay by race already discussed (Price-Williams & Ramirez, 1974; Strickland, 1972). Whereas Watts and colleagues' (2018) sample was only 16% Black, the present sample was almost entirely Black and Bowersox's (2013) sample was over 40% Black. In any case, the results of all three studies are in agreement that low-SES children's *choices* to delay gratification may forecast their likelihood for success in school.

Limitations and Future Directions for Research

The present study has several limitations. Notably, I did not measure the proposed alternative mechanisms other than self-control which may affect children's choice delay performance. On a related note, given the racial homogeneity of the present sample, I am unable to ascertain whether these results are unique to Black and/or Latino/a children from low-income families. As described earlier, processes influencing children's delay of gratification tendencies may differ between demographic groups. Future research should investigate the extent to which children of different races are more or less sensitive to experimental manipulations of contextual cues affecting, for example, expectancy for later rewards. Should contextual cues have differential effects across demographic groups, then the meaning of delay of gratification measures – and perhaps also delay of gratification as a construct – may also differ across groups.

Namely, for groups of children who are relatively unaffected by contextual manipulations, choice delay of gratification behaviors may more closely reflect self-regulatory ability than for groups of children for whom delay of gratification behaviors are highly affected by context.

The results of the present study should not be expected to extend to other measures of delay of gratification. The positive relations between sustained delay of gratification and other outcomes are robust, including in the long-term (Casey et al., 2011; Evans & Rosenbaum, 2008; Moffit et al., 2011) and including in low-SES samples (Watts et al., 2018). There is a dramatic theoretical difference between choice delay of gratification and other forms of delay of gratification; some (e.g., “gift delay,” LiGrining, 2007) do not incorporate concrete rewards given to children contingent on their task performance. Many of the processes, and reward expectancy in particular, may therefore be less confounded in children’s task performance on said tasks. More explicit examination into the skills, biases, experiences, and preferences underlying choice delay – and comparison between those choice and other delay of gratification tasks – has been called for elsewhere (Beck et al., 2001). Such efforts seem well-warranted given the numerous processes speculated to support delay of gratification and may serve to clarify constructs which incorporate both affective and self-regulatory processes (Schwarz et al., 1983).

As with most correlational studies, the present study has a weak warrant for interpreting the observed relations as causal. The numerous alternative, unmeasured processes mentioned earlier which may explain delay of gratification all represent potential confounds in relations between choices to delay gratification and children’s self-regulatory and school-related outcomes. Thus, implications for research and practice, discussed in the next section, must be taken with the fact that the observed relations may not represent causal relations in mind. Future studies incorporating experimental designs may help to clarify whether delaying gratification, as traditionally measured, is indeed as influential as has long been thought.

Implications

Early in delay of gratification research, children's trust in the adult experimenter administering a delay task was acknowledged as a theoretically important byproduct of the immediate social context influencing delay behaviors (Mischel, 1963, 1974; Mischel & Staub, 1965) and, accordingly, trust-building exercises were used in early studies prior to administering sustain delay tasks (e.g., Mischel & Ebbesen, 1970). Such practices may be prudent in light of experimental confirmation of reward expectancy as a confound in delay performance (Kidd et al., 2013; Lee & Carlson, 2015).

The results of Price-Williams and Ramirez (1974) and Strickland (1972) suggesting lower trust among Black children compared to White children, especially when working with White researchers, may at in part be artifacts of the time in which the studies were conducted. Is it possible that the marked racial tension and discord in the Southwest in the 1970s, which were ostensibly behind the results of Strickland (1972), is still marked enough in the Southeast so as to affect the present study participants' perceptions of White research assistants? Certainly, distrust between children and adults across racial lines is not a desirable outcome from the perspective of social progress. But, while this is one of the first studies suggesting negative relations between a delay of gratification measure and other desirable child outcomes, it is not the first to suggest proclivity toward immediate gratification must always reflect deficient self-regulatory ability, especially in contexts of uncertainty (Lee & Carlson, 2015). These notions highlight the importance of considering social context during the research process, especially from the perspective of historically marginalized groups and when psychological theories are being applied in under-studied contexts.

Theoretical interpretations of delay and its predictive power have motivated and influenced efforts to foster self-control in children in educational settings, particularly in schools serving high proportions of children of color from low-income families (O'Leary & Duby, 1979; Strayhorn, 2002). Promotion of self-control, it is thought, may be protective against the myriad challenges and risks that often lead to academic difficulties and delinquency in these populations (Strayhorn, 2002; Ursache, Blair, & Raver, 2012), and yet, such populations have traditionally been underrepresented in the research on which current theory is based (see Moreton, 2014 for a critical review).

The results of the present study do not necessarily call such efforts into question but may have implications for the implementation and delivery of specific programs. Even if immediate gratification is adaptive for some children based on their personal experiences, it does not align well with what society's dominant institutions, including schools, will recognize as desirable and reward in children in the long-term (Finegood & Blair, 2017). And so, promoting such behaviors in children facing disadvantages in these institutions is not necessarily misguided.

However, framing tendencies and preferences such as delayed gratification as reflecting "character," as is currently often the case (e.g., Elias, 2009), may be ill-advised. Doing so may reinforce implicit biases which regard development of children of color from low-income families as deficient. As these differences likely reflect adaptations, framing these tendencies in terms of "character" or even as "skills" may be unnecessarily stigmatizing and may "go against the grain" of children's development and socialization. Presuming instead that that development is optimized to context may lead to greater appreciation of strengths among historically disadvantaged populations which can in turn be leveraged in the context of intervention (Harris et al., 2009).

Conclusion

A decades-long history of delay of gratification research has suggested that delaying gratification is a strong indicator of a child's likelihood of success later in life (Duckworth, 2011; Mischel, 2014). The present study investigated the psychometric properties of a "choice" delay of gratification measure – which require a choice to delay gratification, but not sustaining said choice – among kindergarten-aged children of color from low-income families. While the measure exhibited good reliability and coherence, children who opted for immediate gratification were, according to all other measures, functioning more adaptively than those to opted to delay gratification. These results point to the need for clarification of constructs underlying choice delay of gratification performance, especially among children who have been historically under-represented in research.

Given that those who tend to prefer immediate gratification have been variously described in the literature as impulsive, impatient, or even "victims of their own vulnerabilities" (Mischel & Ayduk, 2011, p. 83), a critical investigation into alternative explanations for delay of gratification performance is warranted. Not only might this represent a step away from the deficit thinking that has historically dominated research in this area (Banks, McQuater, Ross, & Ward, 1983), it may also provide a more nuanced understanding of the development of children's behavioral tendencies in the face of poverty and racial/ethnic minority status. Ultimately, deeper understanding may point to potential ways to alter children's environment in order to build beliefs, perceptions, and preferences which align, rather than compete, with children's latent abilities.

References

Anderson, P. (2002). Assessment and development of executive function (EF) during childhood. *Child neuropsychology*, *8*(2), 71-82.

Addessi, E., Bellagamba, F., Delfino, A., De Petrillo, F., Focaroli, V., Macchitella, L., ... & Sbaffi, A. (2014). Waiting by mistake: Symbolic representation of rewards modulates intertemporal choice in capuchin monkeys, preschool children and adult humans. *Cognition*, *130*(3), 428-441.

Banks, W. C., McQuater, G. V., Ross, J. A., & Ward, W. E. (1983). Delayed gratification in blacks: A critical review. *Journal of Black Psychology*, *9*(2), 43-56.

Beck, D. M., Schaefer, C., Pang, K., & Carlson, S. M. (2011). Executive function in preschool children: Test–retest reliability. *Journal of Cognition and Development*, *12*(2), 169-193.

Bellagamba, F., Addessi, E., Focaroli, V., Pecora, G., Maggiorelli, V., Pace, B., & Paglieri, F. (2015). False belief understanding and "cool" inhibitory control in 3-and 4-years-old Italian children. *Frontiers in Psychology*, *6*, 872.

Blair, C., & Raver, C. C. (2012). Child development in the context of adversity: experiential canalization of brain and behavior. *American Psychologist*, *67*(4), 309.

Bowersox, C. M. (2013). *Construct Validity of Affective and Cognitive Inhibitory Control Tasks for Preschoolers* (Doctoral Dissertation). Retrieved from: https://etda.libraries.psu.edu/files/final_submissions/8589

Bronson, M. B., Goodson, B. D., Layzer, J. I., & Love, J. M. (1990). Child behavior rating scale. *Cambridge: MA: Abt Associates*.

Casey, B. J., Somerville, L. H., Gotlib, I. H., Ayduk, O., Franklin, N. T., Askren, M. K., ... & Glover, G. (2011). Behavioral and neural correlates of delay of gratification 40 years later. *Proceedings of the National Academy of Sciences*, *108*(36), 14998-15003.

Cameron, C. E., Brock, L. L., Murrah, W. M., Bell, L. H., Worzalla, S. L., Grissmer, D., & Morrison, F. J. (2012). Fine motor skills and executive function both contribute to kindergarten achievement. *Child development*, *83*(4), 1229-1244.

Cameron Ponitz, C., McClelland, M. M., Matthews, J. S., & Morrison, F. J. (2009). A structured observation of behavioral self-regulation and its contribution to kindergarten outcomes. *Developmental psychology*, *45*(3), 605.

Cheung, G. W and Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233-255.

Crosby, J. W. (2011). Test Review: FM Gresham & SN Elliott Social Skills Improvement System Rating Scales. Minneapolis, MN: NCS Pearson, 2008. *Journal of Psychoeducational Assessment*, *29*(3), 292-296.

Davis, E. P., Bruce, J., & Gunnar, M. R. (2002). The anterior attention network: Associations with temperament and neuroendocrine activity in 6-year-old children. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology*, *40*(1), 43-56.

Doromal, J. B., Cottone, E. A., & Kim, H. (2017). Preliminary Validation of the Teacher-Rated DESSA in a Low-Income, Kindergarten Sample. *Journal of Psychoeducational Assessment*, 0734282917731460.

Duckworth, A. L. (2011). The significance of self-control. *Proceedings of the National Academy of Sciences*, *108*(7), 2639-2640.

Duckworth, A. L., & Kern, M. L. (2011). A meta-analysis of the convergent validity of self-control measures. *Journal of Research in Personality*, *45*(3), 259-268.

Duckworth, A. L., Tsukayama, E., & Kirby, T. A. (2013). Is it really self-control? Examining the predictive power of the delay of gratification task. *Personality and Social Psychology Bulletin*, *39*(7), 843-855.

Duran, C. A. K., Cottone, E. A., Ruzek, E., Mashburn, A., & Grissmer, D. W. (under revision). Family stress processes and children's self-regulation. *Child Development*

Elias, M. J. (2009). Social-emotional and character development and academics as a dual focus of educational policy. *Educational Policy*, *23*(6), 831-846.

Ellis, B. J., Bianchi, J., Griskevicius, V., & Frankenhuis, W. E. (2017). Beyond risk and protective factors: An adaptation-based approach to resilience. *Perspectives on Psychological Science*, *12*(4), 561-587.

Evans, G. W. (2004). The environment of childhood poverty. *American psychologist*, *59*(2), 77.

Evans, G. W., Gonnella, C., Marcynyszyn, L. A., Gentile, L., & Salpekar, N. (2005). The role of chaos in poverty and children's socioemotional adjustment. *Psychological science*, *16*(7), 560-565.

Evans, G. W., & Rosenbaum, J. (2008). Self-regulation and the income-achievement gap. *Early Childhood Research Quarterly*, *23*(4), 504-514.

Finegood, E. D. & Blair, C. (2017). Poverty, Parent Stress, and Emerging Executive Functions in Young Children. In Deater-Deckar, K. & Panneton, R. (Eds.) *Parental Stress and Early Child Development: Adaptive and Maladaptive Outcomes.* Springer International Publishing AG, Switzerland

Fishbach, A., & Converse, B. A. (2010). Identifying and battling temptation. *Handbook of self-regulation: Research, theory and applications*, *2*, 244-260.

Freire, E., Gormana, B., & Wessman, A. E. (1980). Temporal span, delay of gratification, and children's socioeconomic status. *The Journal of Genetic Psychology, 137*(2), 247-255.

Göllner, L. M., Ballhausen, N., Kliegel, M., & Forstmeier, S. (2018). Delay of gratification, delay discounting and their associations with age, episodic future thinking, and future time perspective. *Frontiers in psychology*, *8*, 2304.

Gresham, F., & Elliott, S. N. (2008). *Social Skills Improvement System (SSIS) Rating Scales.* Minneapolis, MN: NCS Pearson

Hackman, D. A., & Farah, M. J. (2009). Socioeconomic status and the developing brain. *Trends in cognitive sciences*, *13*(2), 65-73.

Hoff, E., & Tian, C. (2005). Socioeconomic status and cultural influences on language. *Journal of communication Disorders*, *38*(4), 271-278.

Hongwanishkul, D., Happaney, K. R., Lee, W. S., & Zelazo, P. D. (2005). Assessment of hot and cool executive function in young children: Age-related changes and individual differences. *Developmental neuropsychology*, *28*(2), 617-644.

Kidd, C., Palmeri, H., & Aslin, R. N. (2013). Rational snacking: Young children's decision-making on the marshmallow task is moderated by beliefs about environmental reliability. *Cognition*, *126*(1), 109-114.

Kim, P., Evans, G. W., Angstadt, M., Ho, S. S., Sripada, C. S., Swain, J. E., ... & Phan, K. L. (2013). Effects of childhood poverty and chronic stress on emotion regulatory brain function in adulthood. *Proceedings of the National Academy of Sciences*, *110*(46), 18442-18447.

Kuntsi, J., Stevenson, J., Oosterlaan, J., & Sonuga-Barke, E. J. (2001). Test-retest reliability of a new delay aversion task and executive function measures. *British Journal of Developmental Psychology, 19*(3), 339-348.

Lee, W. S., & Carlson, S. M. (2015). Knowing when to be “rational”: Flexible economic decision making and executive function in preschool children. *Child Development*, *86*(5), 1434-1448.

Li-Grining, C. P. (2007). Effortful control among low-income preschoolers in three cities: Stability, change, and individual differences. *Developmental Psychology, 43*(1), 208.

Lindstrom, D. R., & Shipman, V. C. (1972). Mischel Technique; Technical Report 12. Disadvantaged Children and Their First School Experiences. *ETS-Head Start Longitudinal Study. Technical Report Series.* Retrieved from: https://eric.ed.gov/?id=ED081825

Lupien, S. J., McEwen, B. S., Gunnar, M. R., & Heim, C. (2009). Effects of stress throughout the lifespan on the brain, behaviour and cognition. *Nature Reviews Neuroscience, 10*(6), 434-445.

Mahrer, A. R. (1956). The role of expectancy in delayed reinforcement. *Journal of Experimental Psychology, 52*(2), 101.

McClelland, M. M., & Morrison, F. J. (2003). The emergence of learning-related social skills in preschool children. *Early Childhood Research Quarterly*, *18*(2), 206-224.

Michaelson, L., de la Vega, A., Chatham, C., & Munakata, Y. (2013). Delaying gratification depends on social trust. *Frontiers in psychology, 4*, 355.

Michaelson, L. E., & Munakata, Y. (2016). Trust matters: Seeing how an adult treats another person influences preschoolers' willingness to delay gratification. *Developmental Science, 19*(6), 1011-1019.

Mischel, W. (2014). *Marshmallow Test.* Random House.

Mischel, W., & Ayduk, O. (2011). Willpower in a cognitive-affective processing system. In Vohs, K. D., & Baumeister, R. F. (Eds.). (2011). *Handbook of self-regulation: Research, theory, and applications.* Guilford Press. 83-105.

Mischel, W., & Baker, N. (1975). Cognitive appraisals and transformations in delay behavior. *Journal of Personality and Social Psychology, 31*(2), 254.

Mischel, W., & Ebbesen, E. B. (1970). Attention in delay of gratification. *Journal of Personality and Social Psychology, 16*(2), 329.

Mischel, W., Ebbesen, E. B., & Zeiss, A. R. (1973). Selective attention to the self: Situational and dispositional determinants. *Journal of Personality and Social Psychology, 27*(1), 129.

Mischel, W., Shoda, Y., & Rodriguez, M. I. (1989). Delay of gratification in children. *Science, 244*(4907), 933-938.

Mischel, W., & Staub, E. (1965). Effects of expectancy on working and waiting for larger reward. *Journal of Personality and Social Psychology, 2*(5), 625.

Mittal, R., Russell, B. S., Britner, P. A., & Peake, P. K. (2013). Delay of gratification in two-and three-year-olds: Associations with attachment, personality, and temperament. *Journal of Child and Family Studies, 22*(4), 479-489.

Moreton, B. (2014). S'More Inequality The Neoliberal Marshmallow and the Corporate Reform of Education. *Social Text, 32*(3 120), 29-48.

Moore, K. A., Vandivere, S., & Ehrle, J. (2000). Turbulence and Child Well-Being. *New Federalism: National Survey of America's Families, Series B*, No. B-16. Assessing the New Federalism: An Urban Institute Program To Assess Changing Social Policies.

O'Leary, S. G., & Dubey, D. R. (1979). Applications of self-control procedures by children: a review. *Journal of Applied Behavior Analysis, 12*(3), 449-465.

Paglieri, F., Addessi, E., Sbaffi, A., Tasselli, M. I., & Delfino, A. (2015). Is it patience or motivation? On motivational confounds in intertemporal choice tasks. *Journal of the experimental analysis of behavior*, *103*(1), 196-217.

Prencipe, A., & Zelazo, P. D. (2005). Development of affective decision making for self and other evidence for the integration of first-and third-person perspectives. *Psychological Science, 16*(7), 501-505.

Price-Williams, D. R., & Ramirez III, M. (1974). Ethnic differences in delay of gratification. *The Journal of Social Psychology, 93*(1), 23-30.

Raver, C. C., Blair, C., Willoughby, M., & The Family Life Project Key Investigators. (2013). Poverty as a predictor of 4-year-olds' executive function: New perspectives on models of differential susceptibility. *Developmental Psychology*, *49*(2), 292-304.

Reynolds, B., & Schiffbauer, R. (2005). Delay of gratification and delay discounting: A unifying feedback model of delay-related impulsive behavior. *The Psychological Record*, *55*(3), 439-460.

Rimm-Kaufman (2005). Survey of Early School Adjustment. Unpublished measure, University of Virginia, Charlottesville, VA.

Ross, L. T., & Hill, E. M. (2002). Childhood unpredictability, schemas for unpredictability, and risk taking. *Social Behavior and Personality: an international journal*, *30*(5), 453-473.

Sarsour, K., Sheridan, M., Jutte, D., Nuru-Jeter, A., Hinshaw, S., & Boyce, W. T. (2011). Family socioeconomic status and child executive functions: The roles of language, home environment, and single parenthood. *Journal of the International Neuropsychological Society*, *17*(1), 120-132.

Schmeichel, B. J. & Baumeister, R. F. (2004). In Vohs, K. D., & Baumeister, R. F. (Eds.). (2004). *Handbook of self-regulation: Research, theory, and applications.* Guilford Press. pp. 84 – 98.

Schwarz, J. C., Schrager, J. B., & Lyons, A. E. (1983). Delay of gratification by preschoolers: Evidence for the validity of the choice paradigm. *Child Development*, 620-625.

Shoda, Y., Mischel, W., & Peake, P. K. (1990). Predicting adolescent cognitive and self-regulatory competencies from preschool delay of gratification: Identifying diagnostic conditions. *Developmental psychology*, *26*(6), 978.

Spann, C. A. (2014). *The Role of Executive Functions And Maternal Characteristics In Externalizing Problems Among Young Children* (Doctoral Dissertation). Retrieved from: https://uta-ir.tdl.org/uta-ir/bitstream/handle/10106/24450/Spann_uta_2502M_12658.pdf?sequence=1&isAllowed=y

Strayhorn, J. M. (2002). Self-control: Toward systematic training programs. *Journal of the American Academy of Child & Adolescent Psychiatry*, *41*(1), 17-27.

Strickland, B. R. (1972). Delay of gratification as a function of race of the experimenter. *Journal of personality and Social Psychology*, *22*(1), 108.

Sturge-Apple, M. L., Suor, J. H., Davies, P. T., Cicchetti, D., Skibo, M. A., & Rogosch, F. A. (2016). Vagal tone and children's delay of gratification: differential sensitivity in resource-poor and resource-rich environments. *Psychological science*, *27*(6), 885-893.

Thompson, C., Barresi, J., & Moore, C. (1997). The development of future-oriented prudence and altruism in preschoolers. *Cognitive Development*, *12*(2), 199-212.

Toner, I. J., Holstein, R. B., & Hetherington, E. M. (1977). Reflection-impulsivity and self-control in preschool children. *Child Development*, 239-245.

Tynan, S. A. (2014). Inhibitory Control and Classroom Behavior in Kindergarten Children (Doctoral Dissertation). *Dissertations.* 82. Retrieved from: http://commons.lib.jmu.edu/diss201019/82

Ursache, A., Blair, C., & Raver, C. C. (2012). The promotion of self-regulation as a means of enhancing school readiness and early achievement in children at risk for school failure. *Child Development Perspectives*, *6*(2), 122-128.

Watts, T. W., Duncan, G. J., & Quan, H. (2018). Revisiting the Marshmallow Test: A Conceptual Replication Investigating Links Between Early Delay of Gratification and Later Outcomes. *Psychological science,* 0956797618761661.

Wilson, A. C., Lengua, L. J., Tininenko, J., Taylor, A., & Trancik, A. (2009). Physiological profiles during delay of gratification: Associations with emotionality, self-regulation, and adjustment problems. *Journal of Applied Developmental Psychology*, *30*(6), 780-790.

Wirth, R. J. and Edwards, M. C. (2007). Item factor analysis: Current approaches and future directions. *Psychological Methods*, *12*(1), 58-79. doi:10.1037/1082-989X.12.1.58.

Wulfert, E., Block, J. A., Santa Ana, E., Rodriguez, M. L., & Colsman, M. (2002). Delay of gratification: Impulsive choices and problem behaviors in early and late adolescence. *Journal of personality*, *70*(4), 533-552.

Yen (1984)

Zenisky, A. L., Hambleton, R. K., & Sireci, S. G. (2001). Effects of Local Item Dependence on the Validity of IRT Item, Test, and Ability Statistics. MCAT Monograph. Accessed July 11, 2018: https://files.eric.ed.gov/fulltext/ED462426.pdf

Tables

Table 1. Univariate Descriptive Statistics

	n	%missing	Mean	SD	Min	Max
Demographics and Home Environment						
FRL	318	14.3%	0.96	0.2	0	1
OPA	327	11.9%	0.81	0.39	0	1
Single Adult in home	326	12.1%	0.54	0.5	0	1
Child is Black	347	6.5%	0.88	0.32	0	1
Child is Male	371	0.0%	0.47	0.5	0	1
Age at Choice Delay Test	354	4.6%	5.42	0.33	4.50	6.17
Mom under 21 at birth	321	13.5%	0.32	0.47	0	1
Mother's Age at Birth	321	13.5%	24.03	5.18	14.59	40.67
Choice Delay Performance						
Item 1 - 1:6 stickers	354	4.6%	0.49	0.50	0	1
Item 2 - 1:2 candies	354	4.6%	0.47	0.50	0	1
Item 3 - 1:4 candies	354	4.6%	0.57	0.50	0	1
Item 4 - 1:2 stickers	354	4.6%	0.53	0.50	0	1
Item 5 - 1:4 stickers	353	4.9%	0.55	0.50	0	1
Item 6 - 1:6 candies	353	4.9%	0.54	0.50	0	1
Choice Delay Total Score	354	4.6%	3.14	2.21	0	6
Choice Delay > 3	354	4.6%	0.49	0.50	0	1
Choice Delay > 2	354	4.6%	0.60	0.49	0	1
Concurrent Self-Regulatory Measures						
Self-Management (DESSA)	313	15.6%	3.61	0.79	1.36	5.00
Responsible Decision-Making (DESSA)	313	15.6%	3.64	0.76	1.13	5.00
Executive Function (HTKS)	354	4.6%	28.79	27.13	0	88
School-Related Outcomes						
Difficulties Transitioning to Kindergarten	310	16.4%	2.06	0.93	1.00	5.00
Self-Control (SSIS) – end of K	318	14.3%	2.90	0.76	1.00	4.00
Self-Regulation (CBRS) – end of K	318	14.3%	3.01	0.69	1.00	4.00
Externalizing (SSIS) – end of K	318	14.3%	1.71	0.62	1.00	3.75
Hyperactivity (SSIS) – end of K	318	14.3%	1.94	0.67	1.00	3.71
WJ - Letter-Word Identification – beg of K	353	4.9%	11.75	5.21	1	39
WJ - Applied Problems – beg of K	353	4.9%	12.23	3.57	0	23
WJ - Academic Knowledge – beg of K	353	4.9%	26.97	4.97	9	41
Difficulties Transitioning to 1st Gr.	222	40.2%	2.33	1.02	1.00	5.00
Self-Control (SSIS) – end of 1st Gr.	216	41.8%	2.78	0.80	1.00	4.00
Self-Regulation (CBRS) – end of 1st Gr.	216	41.8%	2.94	0.73	1.10	4.00
Externalizing (SSIS) – end of 1st Gr.	215	42.0%	1.83	0.66	1.00	4.00
Hyperactivity (SSIS) – end of 1st Gr.	215	42.0%	2.05	0.74	1.00	4.43
WJ - Letter-Word Identification – end of 1st Gr.	292	21.3%	23.38	7.84	6	56
WJ - Applied Problems – end of 1st Gr.	293	21.0%	17.96	3.87	6	36
WJ - Academic Knowledge – end of 1st Gr.	293	21.0%	31.66	4.19	19	44
WJ - Letter-Word Identification – end of 2nd Gr.	250	32.8%	34.28	8.05	14	54

WJ - Applied Problems – end of 2nd Gr.	250	32.8%	22.85	4.08	11	39
WJ - Academic Knowledge – end of 2nd Gr.	250	32.8%	35.37	4	25	45
Other Covariates						
Assessed in Summer	354	4.6%	0.51	0.5	0	1
Naming Vocabulary	353	4.9%	118.10	16.69	64	167
School 2	371	0.0%	0.27	0.45	0	1
School 3	371	0.0%	0.32	0.47	0	1
School 4	371	0.0%	0.06	0.25	0	1
Cohort 2	371	0.0%	0.28	0.45	0	1
Cohort 3	371	0.0%	0.30	0.46	0	1
Offered Treatment	226	0.0%	0.61	0.49	0	1

Table 2. Correlations between Demographic and Control Variables and Choice Delay Performance

	Item 1 - 1:6 stickers	Item 2 - 1:2 candies	Item 3 - 1:4 candies	Item 4 - 1:2 stickers	Item 5 - 1:4 stickers	Item 6 - 1:6 candies	Composite (Sum) Choice Delay	Choice Delay > 3	Choice Delay > 2
FRL	-0.04	-0.01	-0.03	0.01	-0.01	-0.01	-0.02	0	-0.03
OPA	-0.05	-0.02	-0.14 *	-0.01	0.01	-0.02	-0.05	-0.09	-0.06
Single Adult	-0.08	0	0.01	0.04	0.02	0	0	-0.04	0
Black	-0.06	-0.02	0.02	0	0.03	0.04	0	-0.02	0.03
Male	0.01	-0.05	-0.03	0.01	-0.03	-0.08	-0.04	-0.03	-0.02
Age at Test	-0.06	0	0.06	-0.04	-0.06	0.05	-0.01	-0.02	0.02
Mom Under 21	-0.09	-0.11 +	-0.04	-0.1 +	-0.04	-0.06	-0.1 +	-0.06	-0.06
Mom Age	0.03	0.07	0.05	0.09	0.05	0.1 +	0.09	0.03	0.03
School1	-0.05	0	0.04	0.01	0.11 *	0.05	0.04	0.02	0.05
School2	-0.07	-0.03	-0.04	-0.1 +	-0.14 *	-0.11 *	-0.11 *	-0.1 +	-0.13 *
School3	0.09 +	0.02	-0.05	0	0	-0.01	0.01	0.03	0.01
School4	0.05	0.04	0.1 +	0.14 *	0.04	0.11 *	0.11 *	0.09 +	0.13 *
C1	0.05	0.03	0.04	0.04	0.04	0.09 +	0.07	0.04	0.03
C2	0.04	0.02	0	0.06	0.06	-0.04	0.03	0.06	0.03
C3	-0.1 +	-0.05	-0.05	-0.1 +	-0.11 *	-0.06	-0.1 +	-0.11 *	-0.07
Treatment	0.09	0.02	0	0.04	0.04	0.09	0.06	0.02	0.01
Nam. Vocab	-0.04	-0.14 *	-0.11 *	-0.11 *	-0.06	0	-0.11 +	-0.11 *	-0.11 +
Summer Assess.	0.2 ***	0.06	0.12 *	0.17 ***	0.12 *	0.1 +	0.17 ***	0.18 ***	0.12 *

Table 3. Categorical Confirmatory Factor Analysis (CCFA) Fit Parameters: Comparing an Unconstrained and Rasch Model

	Fit Parameters Across 20 Imputed Data Sets	
	Unconstrained IFA Model	Rasch (Constrained) Model
# Free Parameters	12	6
df	9	15
CFI		
Mean	0.995	0.931
SD	0.001	0.004
Min	0.993	0.917
Max	0.997	0.936
TLI		
Mean	0.992	0.931
SD	0.002	0.004
Min	0.989	0.917
Max	0.995	0.936

Table 4. Item Factor Analysis (CCFA) Parameters

	CCFA Factor Loadings		CCFA Item Thresholds		IRT Parameter Equivalents		Tetrachoric Correlations				
	Est. (SE)		Est. (SE)		Disc.	Diff.	Item 1	Item 2	Item 3	Item 4	Item 5
Item 1 - 1:6 stickers	0.68	***	0.03		0.92	0.04					
	(0.06)		(0.07)								
Item 2 - 1:2 candies	0.77	***	0.09		1.22	0.12	0.61				
	(0.04)		(0.07)								
Item 3 - 1:4 candies	0.87	***	-0.17	*	1.80	-0.19	0.55	0.69			
	(0.03)		(0.07)								
Item 4 - 1:2 stickers	0.84	***	-0.06		1.55	-0.07	0.60	0.61	0.72		
	(0.04)		(0.07)								
Item 5 - 1:4 stickers	0.83	***	-0.11	+	1.51	-0.13	0.56	0.64	0.68	0.73	
	(0.04)		(0.07)								
Item 6 - 1:6 candies	0.85	***	-0.09		1.59	-0.11	0.47	0.61	0.78	0.70	0.72
	(0.04)		(0.07)								

Table 5. Frequencies for the 20 Most Frequently Observed Response Patterns

Rank	Response Pattern	Frequency
1	111111	71
2	000000	67
3	100000	16
4	011111	15
5	101111	14
6	001111	12
7	111011	10
8	000001	9
9	111101	8
10	001000	7
11	111110	5
12	110111	5
13	101110	5
14	001001	5
15	000010	5
16	101101	4
17	100110	4
18	011000	4
19	010000	4
20	001101	4

Notes:
"Rank" refers to the rank-ordering of response patterns from most to least frequent. Each "1" in the response patterns indicates a choice to delay gratification on the item corresponding to the digit's position. Correspondingly, each "0" indicates a choice for immediate gratification. For example, "001111" corresponds to a choice to delay on items 3 through 6 and for immediate gratification on items 1 and 2. Response pattern frequencies in this table are derived from the *observed* sample (using listwise deletion).

Table 6. Concurrent Correlations between Choice Delay Performance (Four Scoring Schemes) and Teacher-Reports of Responsible Decision-Making and Self-Management and Direct Measure of Executive Function (HTKS)

	Zero-Order Correlations				Partial Correlations			
	Latent Factor	Composite Sum	High vs. Low	First Item Only	Latent Factor	Composite Sum	High vs. Low	First Item Only
Executive Function (HTKS)	-0.21 ***	-0.19 ***	-0.23 **	-0.17 **	-0.19 **	-0.16 **	-0.19 **	-0.14 *
Responsible Decision-Making	-0.16 *	-0.14 *	-0.26 **	-0.15 *	-0.18 *	-0.15 **	-0.27 **	-0.16 *
Self-Management	-0.14 *	-0.12 *	-0.20 **	-0.14 *	-0.16 *	-0.14 **	-0.23 **	-0.16 *

Table 7. Regression Models Predicting Classroom Behaviors

Est.	Difficulties Transitioning to Kindergarten (n = 316)			Problem Behaviors (End of K)			Positive Behaviors (End of K)		
(SE)	HTKS Only	CD Only	CD + HTKS	HTKS Only	CD Only	CD + HTKS	HTKS Only	CD Only	CD + HTKS
Covariates	X	X	X	X	X	X	X	X	X
CD (Latent Factor)		0.16 **	0.13 *		0.13 *	0.11 *		-0.18 **	-0.15 **
		(0.06)	(0.05)		(0.06)	(0.06)		(0.06)	(0.06)
EF (HTKS)	-0.23 ***		-0.19 *	-0.14 +		-0.14 *	0.21 ***		0.21 ***
	(0.06)		(0.07)	(0.08)		(0.07)	(0.07)		(0.07)
R-squared	0.20	0.19	0.21	0.11	0.11	0.12	0.16	0.15	0.18

Est.	Difficulties Transitioning to 1st Grade			Problem Behaviors (End of 1st Grade)			Positive Behaviors (End of First Grade)		
(SE)	HTKS Only	CD Only	CD + HTKS	HTKS Only	CD Only	CD + HTKS	HTKS Only	CD Only	CD + HTKS
Covariates	X	X	X	X	X	X	X	X	X
CD (Latent Factor)		0.08	0.07		0.13 *	0.10 +		-0.18 **	-0.14 **
		(0.07)	(0.07)		(0.06)	(0.06)		(0.06)	(0.06)
EF (HTKS)	-0.29 ***		-0.31 ***	-0.21 **		-0.21 ***	0.25 ***		0.25 ***
	(0.06)		(0.06)	(-0.22)		(0.07)	(0.07)		(0.06)
R-squared	0.23	0.16	0.24	0.17	0.16	0.19	0.17	0.15	0.19

Table 8. Regression Models Predicting Academic Achievement

Est. (SE)	Concurrent Achievement (Beginning Of K)		
	HTKS Only	CD Only	CD + HTKS
Covariates	X	X	X
CD (Latent Factor)		-0.10 *	-0.03
		(0.05)	(0.04)
EF (HTKS)	0.37 ***		0.37 ***
	(0.05)		(0.05)
R-squared	0.47	0.38	0.48

Est. (SE)	1st Grade Achievement (Beginning of 1st Grade)			1st Grade Achievement, Controlling for K Achievement		
	HTKS Only	CD Only	CD + HTKS	HTKS Only	CD Only	CD + HTKS
Covariates	X	X	X	X	X	X
CD (Latent Factor)		-0.13 *	-0.08		-0.04	-0.01
		(0.05)	(0.05)		(0.04)	(0.04)
EF (HTKS)	0.31 ***		0.31 ***	0.17 ***		0.16 **
	(0.06)		(0.05)			(0.05)
K Achievement				0.61 ***	0.70 ***	0.62 ***
					(0.04)	(0.05)
R-squared	0.42	0.32	0.43	0.62	0.61	0.63

Est. (SE)	2nd Grade Achievement (Beginning of 2nd Grade)			2nd Grade Achievement, Controlling for K Achievement		
	HTKS Only	CD Only	CD + HTKS	HTKS Only	CD Only	CD + HTKS
Covariates	X	X	X	X	X	X
CD (Latent Factor)		-0.11 *	-0.04		-0.04	-0.02
		(0.05)	(0.04)		(0.05)	(0.04)
EF (HTKS)	0.32 ***		0.36 ***	0.15 *		0.16 *
	(0.05)		(0.05)	(0.06)		(0.05)
K Achievement				0.5 ***	0.57 ***	0.50 ***
				(0.07)	(0.06)	(0.08)
R-squared	0.37	0.31	0.38	0.50	0.50	0.51

Table 9. Variance Explained and Choice Delay Performance Coefficients in Prediction/Regression Models – Comparison Across Choice Delay Scoring Schemes

Classroom Behaviors			Covs + HTKS Only	CD as Latent Factor	Composite Sum	First Item Only	High vs. Low Delay
Beginning of Kindergarten	Difficulties Transitioning	CD Coef		**0.13 ***	0.06 *	0.26 *	0.21 +
		Outcome R^2	19.5%	**21.3%**	20.9%	21.2%	20.5%
End of Kindergarten	Problem Classroom Behaviors	CD Coef		**0.11 ***	0.05 +	0.17	0.20 +
		Outcome R^2	10.6%	**11.9%**	11.7%	11.3%	11.5%
	Positive Classroom Behaviors	CD Coef		**-0.15 ****	-0.07 *	-0.29 **	-0.28 *
		Outcome R^2	15.6%	**17.9%**	17.6%	17.6%	17.5%
Beginning of 1st Grade	Difficulties Transitioning	CD Coef		0.07	0.03	**0.49 *****	0.17
		Outcome R^2	23.2%	23.5%	23.4%	**28.6 %**	23.6%
End of 1st Grade	Problem Classroom Behaviors	CD Coef		0.10 +	0.05	**0.46 *****	0.22 +
		Outcome R^2	17.4%	18.5%	18.4%	**22.2%**	18.5%
	Positive Classroom Behaviors	CD Coef		-0.14 **	-0.07 *	**-0.46 *****	-0.29 *
		Outcome R^2	17.1%	19.3%	19.2%	**22.0%**	18.9%
Academic Achievement							
		CD Coef		-0.04	-0.01	-0.04	-0.05
Beginning of Kindergarten		Outcome R^2	47.3%	47.6%	47.4%	47.4%	47.4%
Beginning of 1st Grade		CD Coef		-0.03	-0.02	-0.10	-0.08
		Outcome R^2	42.2%	42.4%	42.3%	42.5%	42.4%
Improvement - Beginning of K. to Beginning of 1st Grade		CD Coef		-0.01	-0.01	-0.04	-0.03
		Outcome R^2	62.4%	62.6%	62.5%	62.5%	62.5%
Beginning of 2nd Grade		CD Coef		-0.08	-0.04	-0.17	-0.12
		Outcome R^2	36.9%	37.6%	37.4%	37.6%	37.3%
Improvement - Beginning of K. to Beginning of 2nd Grade		CD Coef		-0.06	-0.03	-0.12	-0.09
		Outcome R^2	50.4%	50.9%	50.8%	50.8%	50.6%

Figures

Figure 1. Choice Delay Composite (Sum) Performance Distribution