Ready for School: Understanding the Assessment and Development of Young Children's Readiness Skills

---

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

---

by

Jaclyn M. Russo, M.A., M.Ed.

August 2019

Curry Programs in Clinical and School Psychology
Curry School of Education
University of Virginia
Charlottesville, Virginia

APPROVAL OF THE DISSERTATION DEFENSE

This dissertation defense ("Ready for School: Understanding the Assessment and Development of Young Children's Readiness Skills") 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. Amanda Williford)

_____
Committee Member (Dr. Jason Downer)

_____
Committee Member (Dr. Virginia Vitiello)

_____
Committee Member (Dr. Daphna Bassok)

_____Date

TABLE OF CONTENTS

ABSTRACT

This dissertation presents three independent studies that focus on two important aspects of school readiness—(1) the assessment of readiness skills both in preschool and at kindergarten entry and, (2) the combined role of teacher-child interactions and neighborhood features in supporting the development of children's school readiness skills with a particular emphasis on their self-regulation skills. Studies 1 and 2 explored the implementation and validity of readiness assessments in both preschool and at kindergarten entry. In study 1, we found evidence for the feasibility and importance in measuring young children's self-regulation skills using multiple measures—a format which has not been before at scale. In study 2, we explored the validity of a widely-used assessment of readiness skills in preschool over the course of the school year, something which has not been examined previously. Our findings from this study illuminated the difficulty of this measure in differentiating amongst discrete readiness skills within and between children in a classroom. Finally, in study 3, we broadened our measurement of factors that influence children's school readiness to include classroom interactions and neighborhood features and specifically focused on a foundational readiness skill, self-regulation. The results of this study highlighted the importance of neighborhood resources to meet basic and daily needs and a child's individual interactions with their teacher for young children's self-regulation development.

Keywords: *school readiness, neighborhoods, teacher-child interactions*

Conceptual Link

Ready for School: Understanding the Assessment and Development of Young Children's Readiness Skills

School readiness is now widely recognized as a multi-dimensional set of interrelated skills that are critical for young children's success in school and life (McClelland, Acock, & Morrison, 2006; Sabol & Pianta, 2012). Self-regulation in particular is an important set of skills that set the foundation for children's learning and engagement in the classroom (McClelland & Cameron, 2012). Given the evidence highlighting the importance of these readiness skills for children's short and long-term outcomes there has been an increase in the measurement of children's readiness skills both before they arrive to kindergarten and at school entry (National Education Goals Panel, 1995; Boivin & Bierman, 2014; Connors-Tadros, 2014; Bulotsky-Shearer & Fantuzzo, 2011; Sabol & Pianta, 2017; Schmitt, Pratt & McClelland, 2014). One of the primary aims of these assessments is to provide teachers with information that supports them in individualizing instructional practices for their students (Regenstein et al., 2017; Snow & Van Hemel, 2008). In order to do this, it is imperative that these assessments be valid, comprehensive and capable of discriminating amongst individual skills (Miller-Bains, Russo, Williford, DeCoster, & Cottone, 2017; Soderberg et al., 2013; Waterman, McDermott, Fantuzzo, & Gadsden, 2012).

Recent work on the measurement of school readiness skills indicates that a disproportionate number of children from families living at or near the poverty line arrive to school substantially below their higher income peers across these key readiness skills—this is known as the readiness gap (Fitzpatrick, McKinnon, Blair & Willoughby, 2014; Janus & Duku, 2007). The readiness gap emerges early, before children arrive to kindergarten and, without early intervention, often widens and becomes more difficult to close as children move through

elementary school (Daily et al., 2010). This has led to a large investment by states into early childhood education, attempting to increase children's school readiness skills, particularly those with risk factors (i.e., experiences at home or in communities) related to school readiness. Research in this area points to the quality of interactions children have with teachers as critical to benefiting from their experiences in the early school setting (Eisenberg, Valiente, & Eggum, 2010; Hamre & Pianta, 2005; O'Connor & McCartney, 2007). This is especially important for self-regulation development as the classroom presents a new environment where children must successfully navigate activities with both adults and their peers (Rimm-Kaufman, Curby, Grimm, Nathanson, & Brock, 2009). In addition, burgeoning research highlights the influence of children's experiences outside of the classroom, in their neighborhoods and communities, as another important factor in the development of young children's self-regulation skills (Blair & Raver, 2012). Thus, taken together, the assessment of young children's readiness skills and examination of their experiences across contexts (i.e., in the classroom and in neighborhoods) appears critical in supporting school readiness and self-regulation development, in particular. A holistic approach to school readiness will help to foster the ideal conditions for young children to grow and thrive in kindergarten and beyond. More about each of these ideas is described below.

School Readiness

Educators and policymakers are invested in ensuring that children arrive to school with the foundational skills that are needed to thrive academically, socially and emotionally in the classroom (Sabol & Pianta, 2017). These foundational skills traverse a range of early learning domains, including language, literacy, math and social-emotional (National Education Goals Panel, 1995; Boivin & Bierman, 2014; Bulotsky-Shearer & Fantuzzo, 2011; Sabol & Pianta, 2017; Schmitt et al., 2014). Although distinct, these skills work together to promote children's

functioning in the classroom both in the short- and long-run (McClelland et al., 2006; Sabol & Pianta, 2012). In particular, children's early self-regulation skills are critical in supporting their ability to engage in learning activities in the classroom (Clements, Sarama, & Germeroth, 2016; McClelland et al., 2007).

Children's Early Self-Regulation Skills Are Foundational For Learning

Although some disagreement exists among researchers as to a precise definition of self-regulation, it is agreed that it includes the ability to temper strong emotions, inhibit thoughts and behaviors in the moment and direct attention to tasks (Campbell et al., 2016). Children experience rapid development of their self-regulatory skills during the preschool and early elementary time period (McClelland & Cameron, 2012). Children's self-regulation skills are foundational for their learning—for example, during a small group literacy lesson, young children need to be able to direct their attention to the instruction given by their teacher and filter out any other distractions in order to successfully participate in and gain the most from the learning activity. Importantly, early self-regulation skills not only support children's learning in the moment, they have also been shown to be predictive of children's academic success later in elementary school. In fact, young children who display higher social-emotional competence in preschool show greater academic success in fifth grade compared to their peers without these early foundational skills (Sabol & Pianta, 2012).

The significance of these early skills for children's success in the classroom and beyond highlights the importance of ensuring that all children arrive to school equipped with these foundational skills. However, not all children arrive to school with the self-regulation skills needed to engage in learning activities in the classroom and a disproportionate number of these children are from families with low socioeconomic status (Fitzpatrick et al., 2014; Janus &

Duku, 2007). Further, children and families from low SES are also more likely to experience stress at home (e.g., higher residential mobility, parental distress) and be embedded within neighborhoods that present additional strain (e.g., low resources, more violence; Bronfenbrenner & Ceci, 1994; Garmezy & Rutter, 1983). The experience of stressors across early environments places an increased load on the cognitive systems responsible for self-regulation development (Blair, Raver, Granger, Mills-Koonce, & Hibel, 2011; Blair, 2010; Sharkey, Tirado-Strayer, Papachristos, & Raver, 2012). Given the differences in early life experiences, there is wide variability between children's foundational readiness skills when they arrive to school, meaning that ECE programs, schools and teachers are in a position that requires them to support children with a range of learning needs.

The Role of Assessments in Children's School Readiness

Given the importance of early readiness skills for later success, the use of readiness assessments both in preschool and at kindergarten entry has gained increased attention from researchers, policy makers, and practitioners (Ackerman & Coley, 2012; Connors-Tadros, 2014). In fact, all 50 states have developed early learning guidelines (ELG) for ECE programs to support children's school readiness, which has led to the use of readiness assessments in preschool to measure children's early skills (Daily et al., 2010). Further, 70% of states implement some form of an entry assessment to assess children's incoming readiness skills at the beginning of kindergarten (Shields et al., 2016). To help guide stakeholders in the selection of readiness assessments, the National Research Council (2008) has provided recommendations for readiness assessments that offer the following guidance: (1) readiness assessments should measure children's skills individually across foundational early learning domains (i.e., cognitive, language, executive functioning, social-emotional and behavioral) and (2) the intended use of the

assessment data should be clearly delineated for teachers, parents and programs (Regenstein et al., 2017; Snow & Van Hemel, 2008).

While assessments of children's early foundational skills have long been used in clinical and research settings (i.e., Behavior Rating Inventory of Executive Function or the Preschool Self-Regulation Assessment; Gioia, Isquith, Guy, & Kenworthy, 2015; Smith-Donald, Raver, Hayes, & Richardson, 2007), creation of assessments for use by teachers at scale is challenging. Readiness assessments need to be feasible for teachers to administer to every student in their class while also maintaining key psychometric properties, such as reliability and validity. In order for these assessments to be useful and informative they need to be measuring what they intend to measure—both at any given time point and over time (Miller-Bains et al., 2017; Soderberg et al., 2013; Waterman et al., 2012). To do this, assessments should be strongly associated with measures of similar skills and also be able to distinguish amongst the different skills. Further, as programs are using readiness assessment data for a variety of purposes (e.g., helping teachers to select instructional supports, progress monitoring of children and programs, targeting of professional development or implementation of new curricula), it is important that programs be clear in their intended use so that the most appropriate assessment for the purpose can be selected (Schilder & Carolan, 2014). Creating readiness assessments that are comprehensive, valid, scalable and capable of fulfilling varied purposes is a challenging task. A critical examination of these measures allows for a deeper understanding of how to continue to improve assessments of readiness skills so that programs, schools and teachers are in the most advantageous position to provide supports early on that meet the unique needs of every child arriving to school—including program or school-level supports such as, targeted professional

development for teachers and classroom-based supports such as, individualized instructional practices.

The Role of Teacher-Child Interactions in Children's School Readiness

Another approach taken to improve children's school readiness involves large investments into early childhood education. Early childhood education research points to the quality of an individual child's interaction with their teacher as critical to benefiting from their experiences in the early school setting (Pianta & Stuhlman, 2004). When children feel close to their teachers and interactions between a child and teacher are characterized by emotional attunement and sensitivity, children are better able to adjust to school (Baker, Grant, & Morlock, 2008; Pianta & Stuhlman, 2004). In particular, dyadic teacher-child interactions support young children's self-regulation development by providing a secure emotional base in the classroom and this is especially important given that early self-regulation skills undergird children's academic learning (Pianta & Stuhlman, 2004). When children and teachers are emotionally attuned and share positive affect, children are less dysregulated (Williford, Whittaker, Vitiello, & Downer, 2013). However, when high levels of conflict characterize interactions between children and teachers, children are less able to regulate their behaviors (Rudasill, 2011). And, for children who reside in under-resourced neighborhoods or neighborhoods with many risk factors, their self-regulation skills are more likely to be underdeveloped compared to their peers who live in more resourced neighborhoods (Blair et al., 2011). In other words, high quality teacher-child interactions are important for all children but are especially important for children who live in neighborhoods that are characterized by less access to resources, less physical order and fewer safety features as these children experience more daily stressors that negatively impact the development of self-regulation skills.

The Role of Neighborhood Experiences in Children's School Readiness

Given the burgeoning research on the importance of early childhood contexts for self-regulation development, there has been a recent shift in early childhood education policy to better understand how features of children's neighborhoods can also bolster or hinder early development. For example, access to resources such as parks or green space where children can play with peers can reduce feelings of stress that negatively impact the cognitive systems responsible for self-regulation development (Flouri, Midouhas, & Joshi, 2014). Similarly, having safe streets (e.g., sidewalks, crosswalks) that are in good physical condition (e.g., no large pieces of abandoned trash or deteriorated safety signs) allows children to play outside or easily walk to the bus stop without experiencing feelings of stress. Importantly, children interact every day with these features of their immediate neighborhoods.

However, most current research (e.g., McCoy, Connors, Morris, Yoshikawa, & Friedman-Krauss, 2015) that measures features of neighborhoods (e.g., safety or physical condition) uses broad indicators, such as, percent of families living in poverty, crime levels or housing quality, which typically are captured at the census tract-level. However, research has found that for families (and, in particular low-income families where time and money to travel can be more challenging) most of their daily activities take place in approximately a .5-mile radius of their home or what is on average a 10-minute walk (Odgers et al., 2009). Therefore, it is important to capture and better understand features of children's proximal or immediate neighborhoods in order to illuminate both resources and risk factors that influence young children's self-regulation development so that supports (e,g., after-school programming, allocation of funds for new sidewalks) can be advocated for in communities.

Three Study Approach

This three-paper dissertation seeks to better understand two important aspects of school readiness—(1) the assessment of readiness skills both in preschool and at kindergarten entry and, (2) the combined role of teacher-child interactions and neighborhood features in supporting the development of children's school readiness skills with a particular emphasis on their self-regulation skills. When schools elect to use readiness assessments are comprehensive and valid and children's interactions with their teacher are characterized by warmth, sensitivity and high levels of support, children are in the best environment to thrive academically, socially and emotionally. Specifically, readiness assessments provide the necessary information to inform teacher's practice and programs can use the information to monitor progress of teachers and programs and target professional development or new curriculum. And, teacher-child dyadic interactions support children's learning in the classroom through the development of self-regulation skills, which allows children to gain the most from learning activities provided by their teacher. Further, high-quality individual teacher-child interactions are particularly important in supporting the development of self-regulation skills for children who reside in under-resourced neighborhoods with less safety features and physical order because they are more likely to experience higher levels of stress outside the classroom compared to their peers who reside in more resourced neighborhoods. Below is a brief description of each of the three studies that critically examine readiness assessments both in preschool and at kindergarten entry and the importance of teacher-child interactions and neighborhood features for young children's self-regulation development.

Study 1: Including Self-Regulation Skills as Part of a Core Component in a Kindergarten Entry Assessment (KEA)

Leveraging data from a pilot kindergarten entry assessment (KEA), in study 1 we examined the use of two measures of self-regulation (teacher-report and direct assessment) and the concurrent links between these measures and children's early math and literacy skills. The sample included 1,864 kindergarten students across 122 classrooms in Virginia. We were interested in the feasibility with which teachers were able to administer two self-regulation measures at scale for every child in their class—one computerized direct assessment and one rating scale. We found that the majority of teachers felt prepared and confident in administering the assessments and they also felt that the amount of time required to do so was appropriate. We were next interested in the associations between the two measures of self-regulation and children's early math and literacy skills. Here we found that both the teacher-report and direct assessment of self-regulation skills were independently associated with children's mathematics and literacy skills at school entry. In addition, the combination of the teacher-report and direct assessment of self-regulation was associated with students' incoming math skills over and above their main effects. This interaction effect indicated that the positive association between direct assessment of self-regulation and math skills became stronger as students were reported by teachers to have lower self-regulation skills. This interaction was not present of children's early literacy skills, indicating an important link between early self-regulation skills and children's early math abilities. In particular, the interaction effect supports recent findings which suggest that self-regulation skills may be especially important for children's early math learning (Clements et al., 2016). Overall, the findings of this study highlight the importance of assessing self-regulation as a separate early learning domain in KEAs especially given the unique association with early math abilities.

Study 2: Examining the Validity of a Widely-Used School Readiness Assessment in Early Childhood: Implications for Teachers and Early Childhood Programs

The purpose of this study was to examine the validity a widely-used performance-based assessment of children's readiness skills in the fall and spring of preschool. In a sample of 1,109 children across 90 classrooms, we compared children's school readiness skills as assessed by teachers using Teaching Strategies GOLD (TS GOLD) to direct assessments administered by independent data collectors. As expected and consistent with prior research, our findings indicated some evidence of convergent validity with other assessments of theoretically similar skills, although correlations were notably small to modest (Miller-Bains et al., 2017). However, we found limited evidence of differentiation both between skills and between children within a classroom using TS GOLD. Importantly, in this study we expanded previous work by examining how well TS GOLD is able to predict growth in children's directly assessed skills in the spring. Results from these analyses again highlighted difficulty with discrete skill differentiation. Overall, results from this study highlight how if comprehensive readiness assessments are not actually distinguishing amongst children's readiness skills then the data is less useful in terms of helping teachers to provide individualized support for children but also in helping early childhood education programs to know where to target supports for teachers or how to modify curriculum to support children's needs.

Study 3: Advancements in Understanding How Children's Experiences in Neighborhoods and Classrooms Contribute to Self-Regulation

Young children's self-regulation skills are foundational for their success in school. However, a disproportionate number of children from families with low socio-economic (SES) status arrive to school with less well-developed self-regulation skills compared to their more

advantaged peers, as these children and families are more likely to experience feelings of stress across multiple environments—which is particularly harmful to the cognitive systems responsible for self-regulation development (Blair et al., 2011). However, once children enter school, they experience a new environment, the classroom, where they will interact with their teacher—an important aspect of early childhood classrooms that is capable of supporting self-regulation development.

In this study, we aimed to examine how the combination of the quality (warm, supportive) of young children's *individual* experiences with their teacher in the classroom with the features of their *proximal* neighborhoods (safety, physical order and positive resources) impacts self-regulation development during preschool. To do this, we implemented a novel neighborhood coding scheme that used Google Street View to virtually "walk" the streets surrounding (within a .5-mile radius) a child's home. We then combined these proximal neighborhood features with the quality of children's individual interactions with their teacher in the classroom to predict children's growth in self-regulation during preschool. We found that for children who experienced lower-quality interactions with their teacher in the classroom and had few resources in their proximal neighborhoods to meet basic and daily needs (i.e., grocery stores, pharmacy), their self-regulation skills did not exhibit growth during the school year. Whereas, for children who also had few resources in their proximal neighborhoods to meet basic and daily needs but experienced *high-quality* interactions with their teacher, their self-regulation skills grew significantly during the preschool year. These results perhaps suggest that while the experience of residing in an under-resourced neighborhood might negatively impact early self-regulation skills, that high quality individual experiences in the classroom have the potential to be protective and to support children's growth in this area.

Contributions of This Three-Paper Dissertation to the Current Literature

In studies 1 and 2 we explored the implementation and validity of readiness assessments in both preschool and at kindergarten entry. In study 1, we found evidence for the feasibility and importance in measuring young children's self-regulation skills using multiple measures—a format which has not been before at scale. Our findings for this study were particularly important in terms of understanding the math abilities of those children who were reported by their teachers to have low self-regulation skills—for these children, the addition of the direct assessment provided a better understanding of early math skills. These results suggest that the use of multiple measures of self-regulation may help to better understand young children's early math skills and this format may be particularly beneficial for children with underdeveloped self-regulation skills at school entry. We further added to the current literature in study 2 by exploring the validity of a widely-used assessment of readiness skills in preschool over the course of the school year, something which has not been examined previously. Our findings from this study illuminated the difficulty of this measure in differentiating amongst discrete readiness skills within and between children in a classroom—a finding that has implications for how teachers and programs use the data gathered from this assessment.

Finally, in study 3, we broadened our measurement of factors that influence children's school readiness to include classroom interactions and neighborhood features and specifically focused on a foundational readiness skill, self-regulation. In this study, we added to the current literature in two ways, (1) through the measurement of features of children's neighborhoods at a more discrete level than has typically been done in the current literature, and (2) through the combination of these proximal neighborhood features with children's individual experiences in the classroom. These two aims advance the literature by providing both a holistic and more

nuanced view of how an individual child's experiences across contexts (i.e., neighborhood and classroom) influence their self-regulation development. Our findings highlight the importance of capturing both positive features and risk factors of proximal neighborhoods as important avenues through which communities and policymakers might provide resources that will help to support young children's early development. In addition, the results of this study highlighted the importance of a child's individual interactions with their teacher as a protective mechanism for children who are more likely to experience stress outside of the classroom. Taken together, all three papers of this dissertation highlight the complexity of measuring the multidimensional nature of young children's school readiness and emphasize the importance of supporting children across all contexts in their lives to ensure that schools, communities and children are ready for school.

References

Ackerman, D. J., & Coley, R. J. (2012). State Pre-K Assessment Policies: Issues and Status. Policy Information Report. *Educational Testing Service*.

Baker, J. A., Grant, S., & Morlock, L. (2008). The teacher-student relationship as a developmental context for children with internalizing or externalizing behavior problems. *School Psychology Quarterly*, *23*(1), 3.

Blair, C. (2010). Stress and the development of self-regulation in context. *Child Development Perspectives*, *4*(3), 181-188.

Blair, C., & Raver, C. C. (2015). School readiness and self-regulation: A developmental psychobiological approach. *Annual Review of Psychology*, *66*, 711-731.

Blair, C., Raver, C. C., Granger, D., Mills-Koonce, R., Hibel, L., & Family Life Project Key Investigators. (2011). Allostasis and allostatic load in the context of poverty in early childhood. *Development and Psychopathology*, *23*(3), 845-857.

Boivin, M., & Bierman, K. L. (2014). School readiness: Introduction to a multifaceted and developmental construct.

Bronfenbrenner, U., & Ceci, S. J. (1994). Nature-nurture reconceptualized in developmental perspective: A bioecological model. *Psychological Review*, *101*(4), 568-586.

Bulotsky-Shearer, R. J., & Fantuzzo, J. W. (2011). Preschool behavior problems in classroom learning situations and literacy outcomes in kindergarten and first grade. *Early Childhood Research Quarterly*, *26*(1), 61-73.

Campbell, S., Denham, S., Howarth, G., Jones, S., Vick Whittaker, J., Williford, W., … Darling-Churchill, K. (2016). Commentary on the review of measures of early childhood social and emotional development: Conceptualization, critique, and recommendations. *Journal*

of Applied Developmental Psychology, 45, 19-41.

Clements, D. H., Sarama, J., & Germeroth, C. (2016). Learning executive function and early mathematics: Directions of causal relations. *Early Childhood Research Quarterly*, *36*, 79-90.

Connors-Tadros, L. (2014). Information and Resources on Developing State Policy on Kindergarten Entry Assessment (KEA). CEELO FastFacts. *Center on Enhancing Early Learning Outcomes*.

Daily, S., Burkhauser, M. & Halle, T. (2010). A review of school readiness practices in the States: Early Learning Guidelines and Assessments. Early Childhood Highlights (Child Trends), 1(3). Retrieved from: http://www.childtrends.org/wp-content/uploads/2013/05/2010-14-SchoolReadinessStates.pdf

Eisenberg, N., Valiente, C., & Eggum, N. (2010). Self-regulation and school readiness. *Early Education and Development, 21,* 681-698. doi: 10.1080/10409289.2010.497451.

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.

Garmezy, N. E., & Rutter, M. E. (1983). Stress, coping, and development in children. In *Seminar on Stress and Coping in Children, 1979, Ctr for Advanced Study in the Behavioral Sciences, Stanford, CA, US*. Johns Hopkins University Press.

Gioia, G., Isquith, P., Guy, S. & Kenworthy, L. (2015). *Behavior Rating Inventory of Executive Function, Second Edition.* Florida: PAR.

Hamre, B. K., & Pianta, R. C. (2005). Can instructional and emotional support in the first-grade classroom make a difference for children at risk of school failure?. *Child*

Development, *76*(5), 949-967.

Janus, M., & Duku, E. (2007). The school entry gap: Socioeconomic, family, and health factors associated with children's school readiness to learn. *Early Education and Development*, *18*(3), 375-403.

Kagan, S. L., Moore, E., & Bredekamp, S. (Eds.). (1998). *Reconsidering Children's Early Development and Learning Toward Common Views and Vocabulary: National Education Goals Panel*. DIANE Publishing.

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.

McClelland, M. M., Acock, A. C., & Morrison, F. J. (2006). The impact of kindergarten learning-related skills on academic trajectories at the end of elementary school. *Early Childhood Research Quarterly*, *21*(4), 471-490.

McClelland, M. & Cameron, C. (2012). Self-regulation in early childhood: Improving conceptual clarity and developing ecologically valid measures. *Child Development Perspectives, 6,* 136-142. doi:10.1111/j.1750-8606.2011.00191.x

McClelland, M., Cameron, C., Connor, C., Farris, C., Jewkes, A. & Morrison, F. (2007). Links between behavioral regulation and preschoolers' literacy, vocabulary, and math skills. *Developmental Psychology, 43,* 947-959. doi: 10.1037/0012-1649.43.4.947

Miller-Bains, K. L., Russo, J. M., Williford, A. P., DeCoster, J., & Cottone, E. A. (2017). Examining the Validity of a Multidimensional Performance-Based Assessment at Kindergarten Entry. *AERA Open*, *3*(2), 2332858417706969.

National Education Goals Panel (1995). Reconsidering children's early development and

learning: toward common views and vocabulary. Retrieved from: http://govinfo.library.unt.edu/negp/reports/child-ea.htm

National Institute for Early Education Research. (2017). *The state of preschool 2016.* New Brunswick, New Jersey: Barnett, S., Friedman-Krauss, A., Weisenfeld, G., Horowitz, M., Kasmin, R., & Squires, J.

O'Connor, E., & McCartney, K. (2007). Examining teacher–child relationships and achievement as part of an ecological model of development. *American Educational Research Journal*, *44*(2), 340-369.

Pianta, R. C., & Stuhlman, M. W. (2004). Teacher-child relationships and children's success in the first years of school. *School Psychology Review*, *33*(3), 444.

Regenstein, E., Connors, M., Romero-Jurado, R., & Weiner, J. (2017). Uses and misuses of kindergarten readiness assessment results. *The Ounce, 6,* 1-48.

Rimm-Kaufman, S. E., Curby, T. W., Grimm, K. J., Nathanson, L., & Brock, L. L. (2009). The contribution of children's self-regulation and classroom quality to children's adaptive behaviors in the kindergarten classroom. *Developmental Psychology*, *45*(4), 958.

Rudasill, K. M. (2011). Child temperament, teacher–child interactions, and teacher–child relationships: A longitudinal investigation from first to third grade. *Early Childhood Research Quarterly*, *26*(2), 147-156.

Sabol, T. J., & Pianta, R. C. (2012). Patterns of school readiness forecast achievement and socioemotional development at the end of elementary school. *Child Development*, *83*(1), 282-299.

Sabol, T. J., & Pianta, R. C. (2017). The state of young children in the United States: School readiness. *Handbook of Early Childhood Development Programs, Practices, and*

Policies, 3.

Schilder, D., & Carolan, M. (2014). State of the States policy snapshot: State early childhood assessment policies. *Center on Enhancing Early Learning Outcomes*.

Schmitt, S. A., Pratt, M. E., & McClelland, M. M. (2014). Examining the validity of behavioral self-regulation tools in predicting preschoolers' academic achievement. *Early Education and Development*, *25*(5), 641-660.

Shields, K. A., Cook, K. D., & Greller, S. (2016). How Kindergarten Entry Assessments Are Used in Public Schools and How They Correlate with Spring Assessments. REL 2017-182. *Regional Educational Laboratory Northeast & Islands*.

Smith-Donald, R., Raver, C. C., Hayes, T., & Richardson, B. (2007). Preliminary construct and concurrent validity of the Preschool Self-regulation Assessment (PSRA) for field-based research. *Early Childhood Research Quarterly*, *22*(2), 173-187.

Snow, C., & Van Hemel, S. (2008). Early Childhood Assessment: Why, What, and How, Committee on Developmental Outcomes and Assessments for Young Children, Board on Children, Youth, and Families, Board on Testing and Assessment, Division of Behavioral and Social Sciences and Education, National Research Council.

Soderberg J., Stull S., Cummings K., Nolen E., McCutchen D., Joseph G. (2013). Inter-rater reliability and concurrent validity study of the Washington Kindergarten Inventory of Developing Skills (WaKIDS). Unpublished report prepared for the State of Washington Office of Superintendent of Public Instruction. Retrieved from http://www.k12.wa.us/WaKIDS/pubdocs/WaKIDS_Report072613.pdf Google Scholar

Waterman, C., McDermott, P. A., Fantuzzo, J. W., & Gadsden, V. L. (2012). The matter of

assessor variance in early childhood education—Or whose score is it anyway? *Early Childhood Research Quarterly*, *27*(1), 46-54.

Williford, A., Whittaker Vick, J., Vitiello, V. & Downer, J. (2013). Children's engagement within the preschool classroom and their development of self-regulation. *Early Education and Development, 24,* 162-187. doi: 10.1080/10409289.2011.628270

Multi-Measure Assessment of Self-Regulation at School Entry: Links to Early Math and Literacy

Jaclyn M. Russo [1]
Amanda P. Williford [1]

[1]University of Virginia, Center for Advanced Study of Teaching and Learning (CASTL)
405 Emmet Street South, Ruffner Hall, Room 221
Charlottesville, VA 22904

Abstract

The present study piloted the use of two measures of self-regulation (teacher-report and direct assessment) at scale in the context of a KEA. The purpose of this study was to examine the links between these measures and students' incoming mathematics and literacy skills. The sample included 1,864 kindergarten students across 122 classrooms. Both the teacher-report and direct assessment of self-regulation skills were independently associated with mathematics and literacy skills at school entry. In addition, the combination of the teacher-report and direct assessment of self-regulation was associated with students' incoming math skills over and above their main effects. This interaction effect indicated that the positive association between the computer direct assessment of self-regulation and math skills became stronger as students were reported by teachers to have lower self-regulation skills. The findings of this study highlight the importance of assessing self-regulation as a separate early learning domain in KEAs.

Keywords: *school readiness, kindergarten entry assessment, self-regulation, direct assessment, teacher report, early childhood, early academic skills, mathematics*

Multi-Measure Assessment of Self-Regulation at School Entry: Links to Early Math and Literacy

Self-regulation skills have been identified as foundational school readiness skills. Research supports a strong link between young students' self-regulation skills and their academic achievement (Blair & Razza, 2007; Eisenberg, Valiente, & Eggum, 2010; Sabol & Pianta, 2012; Shoda, Mischel, & Peake, 1990). Students who arrive to kindergarten with foundational self-regulation skills are more successful in school and life compared to their peers who enter school far behind in these skills (Hair, Halle, Terry-Humen, Lavelle, & Calkins, 2006). Because of this, policymakers and educators have become increasingly interested in assessing students' school readiness across a range of early learning domains through the use of comprehensive kindergarten entry assessments or KEAs (Conners-Tadros, 2014; Daily, Burkhauser, & Halle, 2010; Lee & Burkman, 2002; U.S. Department of Education & DHHS, 2013). The use of KEAs in the United States has expanded nationally in recent years, with the majority of states implementing some form of a KEA (Connors-Tadros, 2014; Shields et al., 2016). Teachers, school leaders, and policymakers use information from KEAs for a variety of purposes, but most often for individualizing instruction to meet students' unique learning needs (Daily et al., 2010; Shields et al., 2016). However, in order to do this it is important that KEAs discretely measure students' incoming readiness skills in order to most appropriately target needed supports.

Given the strong link between early self-regulation skills and students' academic outcomes, it is surprising that very few states measure students' self-regulation skills separately from other learning domains (Child Trends Data Bank, 2015). Lack of explicit measurement of self-regulation skills as a separate early learning domain may be in part due to a lack of measurement tools that are appropriate for use at scale (Halle & Darling-Churchill, 2016). Currently, teacher-report measures are the most appropriate for use at scale in schools (Halle &

Darling-Churchill, 2016). However, in research and clinical settings, self-regulation is often measured using multiple assessment methods. Different methods of assessment each bring with them particular strengths and weaknesses that deserve careful consideration when choosing how to appropriately measure self-regulation as part of a KEA (Reference Withheld).

In the current study, we piloted the assessment of self-regulation through both teacher report and novel, computerized direct assessment as part of a pilot KEA in one southeastern state in the U.S. We explored the extent to which using both methods of assessment would provide valuable information versus using one assessment. And, we examined the associations between the two self-regulation mesures and young students' early literacy and math skills at school entry.

Self- Regulation is a Key Indicator of Kindergarten Readiness

For decades, researchers have highlighted the particular importance of early self-regulation skills for students' later success in school and life (Blair & Razza, 2007; Eisenberg et al., 2010; Sabol & Pianta, 2012; Shoda et al., 1990). Self-regulation is often defined as the ability to control one's own emotions, behaviors and cognitions (Reference Witheld; McClelland & Cameron, 2012; Reference Withheld). More specifically, self-regulation includes skills such as tempering strong emotions, inhibiting a thought or behavior in the moment, and focusing attention to a task (Reference Witheld Halle & Darling-Churchill, 2016). Self-regulation skills have been shown to be dependent upon context (Reference Witheld; McClelland & Cameron, 2012). For example, young children are able to display better self-regulation skills when they are in reliable versus unreliable situations (Kidd, Palmeri, & Aslin, 2013). This finding suggests that children's display of self-regulation, depends, in part, upon the environment where they are expected to perform these skills. This has important implications for early childhood classrooms

because it is a new environment for young students. This new environment—the classroom—requires young students to frequently shift their attention among multiple activities, as well as remain focused on certain tasks, while also appropriately modulating their emotions and inhibiting their behaviors. Being able to do these things in the classroom supports young students' engagement in learning activities.

Engagement in learning activities in the classroom helps to foster young students' interest in the content they are learning which in turn allows them to gain the most from learning activities (Brock, Rimm-Kaufman, Nathanson & Grimm, 2009). Students with well-developed self-regulation skills show greater gains across early academic domains, such as literacy, language, and mathematics, both in kindergarten and beyond (Blair & Razza, 2007; Clements, Sarama, & Germeroth, 2016; McClelland, Geldhof, Cameron, & Wanless, 2015; Neuenschwander, Rothlisberger, Cimeli, & Roebers, 2012; Sabol & Pianta, 2012; Schmitt, Pratt, & McClelland, 2014). Recent research suggests that the link between self-regulation and mathematics may be particularly strong because students with strong self-regulation skills are better able to frequently shift their thinking and inhibit immediate responses—skills that are especially important when completing mathematics problems (Clements et al., 2016; McClelland et al., 2015; Schmitt et al., 2014). Collectively, the current literature underscores the importance of self-regulation skills for students' early learning by highlighting how these skills do not supersede readiness in academic domains, but rather set the foundation required for engagement in academic learning (Blair & Raver, 2015).

Including Self-Regulation in Kindergarten Entry Assessments

There is wide variation across KEAs both in comprehensiveness of skills assessed and assessment method used (Connors-Tadros, 2014; Shields et al., 2016). KEAs typically assess

students' readiness skills broadly across early learning domains (Shields et al., 2016). However, many readiness assessments do a poor job of differentiating amongst individual skills in ways that are useful for kindergarten teachers (Reference Withheld). Given the robust link between self-regulation and academic outcomes, there is a strong rationale for including a discrete assessment of early self-regulation assessment as part of comprehensive KEAs, yet few states measure self-regulation skills of students separately from other learning domains (Blair & Razza, 2007; Child Trends Data Bank, 2015; Connors-Tadros, 2016; Eisenberg et al., 2010; McClelland, Acock, & Morrison, 2006; Shoda et al., 1990). Below, we describe how self-regulation has been assessed in the early education research literature and the potential benefits and drawbacks of these methods.

Benefits and Drawbacks of Currently Available Self-Regulation Measures

There is a long history of assessing young children's self-regulation in the early education research literature with assessment methods including structured observations in laboratory or clinic settings, teacher or adult report, and/or direct assessment (e.g., Reference Witheld; Diamond, Carlson, & Beck, 2005; Reference Withheld; Ponitz, McClelland, Matthews, & Morrison, 2009; Wakschlag et al., 2008; Wanless et al., 2011). Each of these assessment methods offers benefits and drawbacks. Structured observations of young children's self-regulation skills have been used in research and demonstrate good reliability (e.g., Disruptive Behavior Diagnostic Observation Schedule, DB-DOS; Wakschlag et al., 2008). However, no naturalistic observations of self-regulation skills specifically as displayed in the classroom context have been identified (Reference Witheld).

Parent and teacher report of students' self-regulation skills is a commonly used method in research and practice (Halle & Darling-Churchill, 2016; Reference Witheld). Multiple teacher-

report measures exist that have been shown to be both reliable and valid (e.g., Behavior Rating Inventory of Executive Function, Second Edition [Gioia, Isquith, Guy, & Kenworthy, 2015] and Child Behavior Rating Scale [Bronson, Goodson, Layzer, & Love, 1990]). One of the benefits of teacher report of self-regulation includes the time needed to complete the measure—most are quick to complete. Further, teachers spend significant time with their students, making them valuable informants of students' self-regulation skills; however, teacher ratings have been shown to sometimes be more associated with teacher characteristics than students' actual measured skills (Mashburn et al., 2006; Mashburn & Henry, 2004). For example, years of experience and/or expectations of students' skills often influences teachers' ratings of young students' readiness skills (Mashburn et al., 2006; Mashburn & Henry, 2004). In addition, the composition of a teacher's class (e.g., a classroom in which the majority of students are arriving to school with well-developed readiness skills versus a classroom with few students who are ready for school) may influence their ratings of students' readiness skills, such that a teacher may rate a student's skills relative to the other students in the class which may not be accurately reflective of actual skills of individual students (Mashburn & Henry, 2004). And, recent research highlights that when teachers perceive the environment of their school to be favorable, they tend to report fewer initial internalizing and externalizing symptoms in their students (Pas & Bradshaw, 2014). Thus, when considering using teacher report as part of a KEA, it is feasible that teachers could complete a short survey of each student in their classroom. However, the process of completing the same survey on each child could amplify the bias associated with teacher report.

Direct assessments of self-regulation have also been widely used in early childhood research (e.g., Dimensional Change Card Sort Task [DCCS Task; Diamond et al., 2005; Zelazo, 2006]; Head-Toes-Knees-Shoulders [Ponitz et al., 2009]; Preschool Self-Regulation Assessment

[Smith-Donald, Raver, Hayes, & Richardson, 2007]). Recently, some of these direct assessments have been programmed to be administered using computers or tablets (e.g., Social-Emotional Learning, web-based [McKown et al., 2016]; Early Years Toolbox [Howard & Melhuish, 2016] and NIH Toolbox [DCCS Task; Zelazo & Carlson, 2012]). Benefits of direct assessments include that the standard administration reduces bias and allows for comparisons of children within and across classroom (Reference Witheld; Diamond et al., 2005; Schmitt et al, 2014; Wanless et al., 2011). Drawbacks include that direct assessment procedures typically involve removing the child from the classroom and can be time-consuming. For children who may be most likely to display deficits in self-regulation skills, these assesments which occur in more ideal settings could overestimate children's actual capacity to self-regulate in the classroom environment (Reference Witheld). However, this method is also capable of highlighting young students' strengths that might be more difficult for a teacher to observe in the context of the classroom.

In summary, assessing students' self-regulation skills at scale in the context of KEAs requires that the assessment methods produce reliable and valid results and are feasible for teachers to administer to each student in their classroom. Given the information above, teacher report and direct assessment are viable methods for assessing self-regulation as a part of a KEA. It may be beneficial to include both assessment methods because some students may be better able to self-regulate under specific circumstances (e.g., in a one-on-one distraction-free environment), but are less able to do so when in other environments (e.g., centers or circle time in the classroom). In other words, teacher report and direct assessment of students' self-regulation skills may serve to complement one another in certain situations by providing unique information about individual students' self-regulation skills. This may be particularly important

for students with less well-developed early self-regulation skills because their behavior is often more complex and difficult to assess through observation alone. For example, a student might have difficulty following multi-step directions because he has difficulty attending to his teacher during the presentation of instructions, which causes him to miss the instructions altogether, or because he is impulsive and is quick to act upon the beginning of the instructions, missing other important information. In a situation such as this one, multiple assessment methods are advantageous in order to help a teacher both understand this student's behavior and select effective and appropriate supports.

Present Study

The current study piloted the use of two different measures of self-regulation to be used at scale as part of a KEA in one U.S. southeastern state. To our knowledge, this is the first study to examine the assessment of self-regulation using multiple measures within a KEA. We were interested in the following questions: First, would these measures show evidence of basic reliability and validity consistent with prior research using the selected measures? We expected both the teacher-report and the direct assessment measures to exhibit acceptable reliability (i.e., Internal Consistency) as reported in previous research (e.g., McKown et al., 2016; Schmitt et al., 2014). With regard to validity, we expected the two measures to be modestly associated with each other given prior research which shows moderate associations between teacher report and direct assessment (Schmitt et al., 2014). Second, would the two measures of self-regulation be associated with students' incoming math and literacy skills? Here we expected both the teacher report and direct assessment of self-regulation skills to be uniquely and positively associated with students' early mathematics and literacy skills (Clements et al., 2016; McClelland et al., 2015; Schmitt et al., 2014). Finally, we explored whether the interaction between teacher report

and direct assessment of self-regulation skills would be predictive of students' literacy and math skills over and above their main effects. Here we hypothesized that the positive association between students' direct assessment of self-regulation skills and their early academic skills to be stronger for students whom were reported to have low self-regulation skills by their teachers whereas we expected that teacher report would be sufficient for students who were reported by their teachers to display well-developed self-regulation skills.

Methods

Participants

This study utilized archival data from a KEA in a U.S. southeastern state collected in the fall of 2014. Participants were kindergarten students purposely recruited from 44 schools and 122 classrooms to provide a sample that was representative of students across the state (N=1,864 students). Among participating schools, 48.8% qualified for school-wide Title 1 services and 19.1% identified as rural. Child characteristics were diverse, as depicted in Table 1. The majority of participating teachers were Caucasian (85.7%), with a mean age of 41 years (*SD* = 11.09 years), and over 50% had an Master's degree or above. On average, teachers had approximately 14 years of teaching experience (*SD* = 9.5 years).

Procedure

Participating teachers attended a half-day training workshop during the summer months. Teachers received a stipend and lunch as part of their attendance at training. Participating kindergarten teachers were trained to administer the self-regulation assessments by project personnel. The publisher of the mathematics assessment provided an instructor for the training of this assessment. Schools provided separate training for administration of the literacy assessment.

Data collection took place in the fall of the academic year—four to five weeks after the start of the school year. Teachers or another trained personnel administered the direct assessments to students during the school day. Lead classroom teachers completed the teacher report measure of self-regulation. All data were collected online and entered into a secure system by teachers and in accordance with the university's IRB. During active data collection, project staff completed observations of fidelity of assessment administration for 15% of the sample and results indicated good to adequate administration fidelity.

Measures

Self-regulation: Direct assessment. *Social-Emotional Learning, web-based* (SELweb; McKown et al., 2016) is a computer-based, direct assessment of students' social-emotional learning skills. All aspects of the assessment are audio narrated and require few reading skills. The entire SELweb assessment contains five modules: Non-Verbal Awareness, Perspective Taking, Social Problem Solving, Delay of Gratification, and Frustration Tolerance. A confirmatory factor analysis (CFA) was performed taking into account nesting of students in classrooms to examine a hypothesized four-factor model of social-emotional learning (i.e., Social Awareness, Social Meaning, Social Reasoning and Self-Control). The fit of this model to the data was excellent (CFI = 0.96, RMSEA = 0.049 (90% CI 0.04 - 0.06)), and confirmed the four hypothesized factors of social-emotional comprehension (McKown et al., 2016).

As part of this pilot KEA, students only completed three modules within SELweb. Specifically, they completed Social Problem Solving, Delay of Gratification, and Frustration Tolerance, to yield two domains, Social Reasoning (i.e., Social Problem Solving) and Self-Control (i.e., Delay of Gratification and Frustration Tolerance). To keep language consistent, in this study, the Self-Control domain of SELweb is referred to as the Self-Regulation domain. The

present study examined only the modules within the Self-Regulation domain (i.e., Delay of Gratification and Frustration Tolerance modules).

The Delay of Gratification module explores how willing a student is to endure a seemingly boring task to gain the most possible points, while the Frustration Tolerance module assesses a student's ability to inhibit an immediate response while withstanding a mildly frustrating task (McKown et al., 2016). Both modules are scored by creating a summed score of a student's total points. Please refer to McKown et al. (2016) for complete scoring details of SELweb.

SELweb was validated on a large, diverse sample of children in grades K-3 (N=4,462; McKown et al., 2016). The SELweb self-regulation modules demonstrate good convergent validity with alternate measures of self-regulation (i.e., KiTAP, Test of Attentional Performance for Children across two validation studies; McKown et al., 2016). In addition, SELweb is positively and significantly associated with teacher report of social-emotional competencies and negatively associated with teacher-reported problem behaviors (McKown et al., 2016). SELweb is also positively associated with students' early academic skills (i.e., AIMSweb Reading and Math and teacher-reported academic competence on the Social Skills Improvement System rating scale; McKown et al., 2016). In addition, McKown et al. (2016) report acceptable internal consistency of the SELweb social reasoning and self-regulation modules across samples of children (Delay of Gratification α= 0.71 and 0.74; Frustration Tolerance α = 0.77 and 0.92; Social Reasoning α = 0.82 and 0.88). In our sample, the Frustration Tolerance module of the SELweb demonstrated good inter-item reliability (α = 0.85), however, the inter-item reliability for the Delay of Gratification module was weak (α = 0.47). Furthermore, the correlation between the two SELweb self-regulation modules (Frustration Tolerance and Delay of Gratification) was

low (r = 0.13). Given this, the Delay of Gratification module of the SELweb was dropped from further analyses. All analyses examining the relationship of students' self-regulation skills with their academic achievement include only the Frustration Tolerance module.

Self-regulation: Teacher report. Teachers reported on students' self-regulation and social skills using the *Child Behavior Rating Scale* (CBRS; Bronson et al., 1990). The CBRS is an instrument focused on students' behavior with other students and adults in the classroom. The CBRS includes 17 items and yields two broad domains of a student's social-emotional skills: Social Skills and Self-Regulation (Bronson et al., 1990). In the present study, only the 10 items comprising the Self-Regulation domain were examined. Items are assessed using a five-point scale (1= never, 5= always). Some examples of items within the Self-Regulation domain include, "Concentrates when working on a task; is not easily distracted by surrounding activities" and "Observes rules and follows directions without requiring repeated reminders" (Bronson et al., 1990). Students receive an overall self-regulation score by summing the responses for the 10 items.

The CBRS is frequently used in early childhood populations and has been used at kindergarten entry (Cameron Ponitz et al., 2009; Oregon Department of Education). The CBRS demonstrates evidence to support sound construct validity when compared to a direct assessment of self-regulation for a sample of kindergarten students. Students rated as having better developed self-regulatory abilities in the spring of their kindergarten year on the CBRS as rated by their teacher also scored higher on a direct assessment of self-regulation (Cameron Ponitz et al., 2009). Additionally, the CBRS significantly predicts students' achievement scores in math and sound awareness in kindergarten (Cameron Ponitz et al., 2009; McClelland et al., 2007; Schmitt et al., 2014). Furthermore, internal reliability for the self-regulation domain is strong in

recent studies, ranging from α= 0.89-0.96 (Cameron Ponitz et al., 2009; Schmitt et al., 2014). In our sample, the self-regulation domain of the CBRS demonstrated strong inter-item reliability consistent with previous research using the measure (α = 0.97).

Students' early math skills. Teachers administered the *Tools for Early Assessment in Mathematics-Short Form* (TEAM-SF; Weiland et al., 2012) to assess students' early mathematics skills. The TEAM-SF is a 20-item measure of students' number sense and geometric/spatial competencies. Students are directly assessed one-on-one with their teacher or a trained administrator using a standardized protocol. Items assessing number sense include asking students to count to five, count a certain number of objects, subitize objects, and match amounts to numbers. Items assessing geometry include asking students to identify certain shapes, make shapes with straws, and identify certain aspects of shapes (Weiland et al., 2012). Each item is scored as correct or incorrect, with a few items offering partial credit. The final score for students is calculated as a percent correct. In past research, the TEAM-SF has demonstrated good validity for the assessing numeracy and geometry skills of kindergarten students (Weiland et al., 2012). The TEAM-SF has also demonstrated adequate internal consistency, ranging from 0.71- 0.79 (Weiland et al., 2012). In the present study, inter-item reliability of the TEAM-SF was strong (α= 0.94).

Students' early literacy skills. Teachers administered the *Phonological Awareness Literacy Screening-Kindergarten* (PALS-K; Invernizzi, Juel, Swank, & Meier, 2015). PALS-K is a widely used diagnostic assessment at kindergarten entry and is administered one-on-one to students after the first six weeks from the start of the school year. The PALS-K measures students' print knowledge and phonological awareness. More specifically, the PALS-K examines students' rhyme awareness, alphabet knowledge, beginning sound awareness, knowledge of

letter sounds, spelling, concept of word, and word recognition (Invernizzi et al., 2015). Students' scores on each of the subtasks are added together to create a summed score. Across several years, the PALS-K has demonstrated strong test-retest (α= 0.78- 0.95) and inter-rater reliabilities (α= 0.96- 0.99). In addition, concurrent validity with the Standford-9 was strong (r = 0.72; Invernizzi et al., 2015). In the current work, the PALS-K demonstrated sound internal consistency (r = 0.70).

Results

Demographic and descriptive statistics for study participants and predictor and outcome variables are presented in Table 1.

Evidence for Validity

Correlations between the measures of self-regulation and academic achievement were examined as evidence of construct validity and can be found in Table 2. The association between the CBRS and the SELweb was moderate (r = 0.24) and both were significantly and positively associated with students' mathematics and literacy skills.

A regression model predicting students' literacy and mathematics achievement from their teacher-reported and directly-assessed self-regulation was examined. Within this model, the following two associations were examined, (1) the unique associations of the two self-regulation measures with students' academic achievement and, (2) the interaction between the two self-regulation measures with students' academic achievement. Both students' directly-assessed and teacher-reported self-regulation were significant and positively associated with students' mathematics and literacy achievement (see Table 3). Additionally, there was a significant interaction effect indicating that the combination of students' teacher-reported and directly-assessed self-regulation skills was uniquely associated with their mathematics skills over and above the main effects. This interaction effect was not present for students' literacy skills. The

interaction effect is illustrated in Figure 1 and indicated that as teachers reported students to display lower self-regulation skills, the positive association between their directly assessed self-regulation skills and their mathematics skills increased. For students whom teachers indicated as displaying lower self-regulation skills within the classroom, there was a stronger positive association between students' directly assessed self-regulation skills and mathematics skills. In contrast, for students whom teachers reported to display high self-regulation skills, there was a weaker positive association between their directly assessed self-regulation skills and their mathematics skills in the fall of kindergarten.

Discussion

Although many states are developing KEAs to assess students' incoming readiness skills across early learning domains, most do not explicitly measure self-regulation as a separate learning domain, despite evidence highlighting the particular importance of these skills for setting the foundation for learning (Blair & Raver, 2015; Clements et al., 2016; Schmitt et al., 2014). Given the increased awareness of the importance of early self-regulation skills for young students' success in school, we took advantage of recent innovations in direct assessments and piloted the use of two assessments of self-regulation (teacher report and direct assessment) in the context of a KEA. This study replicates the well-established link between self-regulation and academic achievement (Blair & Razza, 2007; Clements et al., 2016; McClelland et al., 2015; Schmitt et al., 2014), and highlights the benefits of using multiple measures of self-regulation at kindergarten entry, especially for students who enter school with less well-developed self-regulation skills.

Validity Findings

Our validity results suggest that both measures of self-regulation are valid when used within the context of a KEA. Teacher report (CBRS) and direct assessment (SELweb Frustration Tolerance) of self-regulation were modestly correlated with one another which was expected given that the measures are not perfectly aligned—the CBRS items assess self-regulation at a more global level within the context of the classroom while the SELweb assesses a specific skill within the broader domain of self-regulation. Although our associations between teacher-reported and directly assessed self-regulation were somewhat weaker than what has been reported in prior research (e.g., Cameron Ponitz et al., 2009, reported an association of $r = 0.29$; and, Schmitt et al., 2014 reported an association of $r = 0.35$), the relationship was modest and in the expected direction offering evidence of validity given that the two measures intentionally assess different aspects of self-regulation skills. Consistent with previous research (Cameron Ponitz et al., 2009; Goldstein et al., 2014; Schmitt et al., 2014), teacher report and direct assessment of self-regulation were significantly and positively associated with students' early mathematics and literacy skills (i.e., students with higher teacher-reported and directly assessed self-regulation skills also displayed higher mathematics and literacy skills). Further, both assessments uniquely predicted students' early academic achievement—students' directly assessed and teacher-reported self-regulation skills were associated with their early mathematics abilities.

We also found that the combination of assessment methods provided a unique association with students' mathematics skills. This finding indicates that the teacher report and direct assessment of self-regulation provide distinctive information about students' self-regulation skills that are important for understanding their early mathematics achievement. Consistent with our hypothesis, this finding was especially important for students whom teachers indicated as

displaying lower self-regulation skills in the classroom at kindergarten entry. In contrast, for students whom teachers reported as displaying well-developed self-regulation skills in the classroom, their directly assessed self-regulation was not as useful in predicting their mathematics achievement. Recent research (Schmitt et al., 2014; Clements & Sarama, 2016) suggests that self-regulation may be especially important for students' early mathematics skills compared to literacy skills, which may explain why this effect was present for students' mathematics but not literacy achievement. This finding helps to illuminate the role that self-regulation skills play in young students' early math skills and the importance of intentionally structuring math activities in the classroom to foster early self-regulation skills.

Implications for the Classroom and Teacher Practice

Information from both assessments could be used to help teachers better understand students' needs in the classroom. For example, some students may have difficulty accessing and using their self-regulation skills in the classroom because they are easily overwhelmed in the moment. However, if these skills were measured using a direct assessment, which is out of the context of the classroom, they may be better able to display some of these skills—highlighting strengths. For these children, an additional measure of self-regulation would be recommended to help their teacher best understand their self-regulation skills and select appropriate strategies to support their learning. Our finding related to young students' early math abilities suggests that not only are self-regulation skills especially important for understanding early math skills but that for those students who display lower self-regulation skills at school entry, the use of two measures can help to inform the best ways to support early learning. For example, early math activities require students to hold information in their mind and inhibit an immediate response in the moment. For students who need support in these skills, teachers can structure activities by

providing visual cues or reminders in the moment to scaffold the development of early self-regulation abilities. When this is done early on in the early childhood classroom setting, young students are well-positioned to develop these skills early and carry them forward through school, ultimately supporting their learning later on.

Limitations

This study was cross-sectional and causality between self-regulation and academic skills cannot be inferred. Additionally, although student and school demographics were varied and diverse, participants in this study were restricted to one geographic area, limiting the applicability of the results to other regions across the U.S. Importantly, the weak reliability of the Delay of Gratification module within the SELweb in our sample should be further explored to better understand how this task functions in samples of young children.

Future Directions

This study provides support for the feasibility of administering multiple measures of self-regulation at scale in schools. As state leaders continue to examine how to structure their KEAs, they should carefully consider the evidence highlighting the importance of explicitly measuring self-regulation as a separate learning domain. In addition, state leaders should consider the benefits of including multiple assessments of self-regulation especially for students who show difficulties in this area upon school entry. It is important to acknowledge that readiness assessments will be most beneficial for students when teachers are able to accurately interpret the data and select and implement appropriate strategies and supports to help close the readiness gap over the kindergarten year. Our study did not examine teachers' ability to interpret their data and/or select resources for students. Future studies should carefully explore how much support

teachers need when interpreting their readiness data in addition to providing direct assistance in selecting supports for students.

Conclusions

We found that the use of multiple measures of self-regulation can serve to complement one another when understanding how self-regulation is linked with students' literacy and mathematics skills. In particular, the use of a direct assessment of self-regulation at scale by teachers is novel because few direct assessments of self-regulation exist that are appropriate for use by teachers at scale. To our knowledge, our study is the first to examine not only the feasibility and validity of multiple measures of self-regulation at scale, but also the links of these measures to students' early mathematics and literacy skills. Importantly, our results indicate that the use of two measures may be especially important for students who display low self-regulation skills in the classroom. These students' behaviors are often complex and difficult to assess, making the use of multiple measures of self-regulation critical in capturing all aspects of their behavior and skills.

References

Aiken, L & West, S. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.

Blair, C. & Raver, C. (2015). School readiness and self-regulation: A developmental psychobiological approach. *Annual Review of Psychology, 66,* 711-731. doi: 10.1146/annurev-psych-010814-015221.

Blair, C. & Razza Peters, R. (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

Brock, L., Rimm-Kaufman, S., Nathanson, L., & Grimm, K. (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,* 337-349. doi: 10.1016/j.ecresq.2009.06.001

Bronson, M., Goodson, B., Layzer, J., & Love, J. (1990). *Child Behavior Rating Scale*. Cambridge, MA: Abt Associates.

Bulotsky-Shearer, R. & Fantuzzo, J. (2011). Preschool behavior problems in classroom learning situations and literacy outcomes in kindergarten and first grade. *Early Childhood Research Quarterly, 26,* 61-73. doi:10.1016/j.ecresq.2010.04.004

Bulotsky-Shearer, R., Fernandez, V., Dominguez, X. & Rouse, H. (2011). Behavior problems in learning activities and social interactions in head start classrooms and early reading, mathematics, and approaches to learning. *School Psychology Review, 40,* 39-56.

Cameron Ponitz, C., McClelland, M., Matthews, J., & Morrison, F. (2009). A structured observation of behavioral self-regulation and its contribution to Kindergarten outcomes. *Developmental Psychology, 45,* 605-619. doi: 10.1037/a0015365

Child Trends Data Bank (2015). *Early school readiness, indicators on children and youth.* Retrieved from: http://www.childtrends.org/wp-content/uploads/2012/10/07_School_Readiness.pdf

Clements, D., Sarama, J. & Germeroth, C. (2016). Learning executive function and early mathematics: Directions of causal relations. *Early Childhood Research Quarterly, 36,* 79-90.

Connors-Tadros, L. (2014). *Information and resources on developing state policy on kindergarten entry assessment (KEA) (CEELO FASTFacts)*. New Brunswick, NJ: Center on Enhancing Early Learning Outcomes.

Daily, S., Burkhauser, M. & Halle, T. (2010). *A review of school readiness practices in the States: Early Learning Guidelines and Assessments*. *Early Childhood Highlights (Child Trends)*. Retrieved from: http://www.childtrends.org/wp-content/uploads/2013/05/2010-14-SchoolReadinessStates.pdf

Diamond, A., Carlson, S., & Beck, D. (2005). Preschool children's performance in task switching on the dimensional change card sort task: Separating the dimensions aids the ability to switch. *Developmental Neuropsychology, 28,* 689-729.

Dusenbury, L., Weissberg, R., Goren, P. & Domitrovich, C. (2014). *State Standards to Advance Social and Emotional Learning: Findings from CASEL's State Scan of Social and Emotional Learning Standards, Preschool through High School*. Retrieved from the Collaborative for Academic, Social, and Emotional Learning website:

http://www.casel.org/state-standards-for-social-and-emotional-learning/

Eisenberg, N., Valiente, C., & Eggum, N. (2010). Self-regulation and school readiness. *Early Education and Development, 21,* 681-698. doi: 10.1080/10409289.2010.497451.

Fitzpatrick, C., McKinnon, R., Blair, C., & Willoughby, M. (2014). Do preschool executive function skills explain the school readiness gap between advantaged and disadvantaged children? *Learning and Instruction, 30,* 25-31. doi: http://dx.doi.org/10.1016/j.learninstruc.2013.11.003

Gioia, G., Isquith, P., Guy, S. & Kenworthy, L. (2015). *Behavior Rating Inventory of Executive Function, Second Edition.* Florida: PAR.

Goldstein, J., Eastwood, M., & Behuniak, P. (2014) Can teacher ratings of students' skills at kindergarten entry predict kindergarten retention? *The Journal of Education Research, 107*(3), 217-229. doi: 10.1080/00220671.2013.788986.

Halle, T. & Darling-Churchill, K. (2016). Review of measures of social and emotional development. *Journal of Applied Developmental Psychology, 45,* 8-18.

Hair, E., Halle, T., Terry-Humen, E., Lavelle, B., & Calkins, J. (2006). Children's school readiness in the ECLS-K: Predictions to academic, health, and social outcomes in first grade. *Early Childhood Research Quarterly, 21,* 431-454. doi: 10.1016/j.ecresq.2006.09.005.

Howard, S. J., & Melhuish, E. (2016). An early years toolbox for assessing early executive function, language, self-Regulation, and social development validity, reliability, and preliminary norms. *Journal of Psychoeducational Assessment*, *35*, 255-275. doi: 10.1177/0734282916633009

Invernizzi, M., Juel, C., Swank, L., & Meier, J. (2014). *PALS Kindergarten Technical Reference*.

Retrieved from: https://pals.virginia.edu/pdfs/rd/tech/K_Tech_Ref_2014_B.pdf

Janus, M. & Duku, E. (2007). The school entry gap: Socioeconomic, family, and health factors associated with children's school readiness to learn. *Early Education and Development, 18,* 375-403.

Kidd, C., Palmeri, H., & Aslin, R. (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.

Kupersmidt, J. B., Stelter, R., & Dodge, K. A. (2011). Development and validation of the social information processing application: A web-based measure of social information processing patterns in elementary school-age boys. *Psychological Assessment, 23*, 834-847. doi:10.1037/a0023621

Lee, V. & Burkman, D. (2002). *Inequality At The Starting Gate : Social Background Differences In Achievement As Children Begin School*. Washington, D.C.: Economic Policy Institute.

Mashburn, A. J., Hamre, B. K., Downer, J. T., & Pianta, R. C. (2006). Teacher and classroom characteristics associated with teachers' ratings of prekindergartners' relationships and behaviors. *Journal of Psychoeducational Assessment*, *24*, 367-380.

Mashburn, A. J., & Henry, G. T. (2004). Assessing school readiness: Validity and bias in preschool and kindergarten teachers' ratings. *Educational Measurement: Issues and Practice*, *23*(4), 16-30.

McClelland, M., Acock, A., & Morrison, F. (2006). The impact of kindergarten learning-related skills on academic trajectories at the end of elementary school. *Early Childhood Research Quarterly, 21,* 471-490. doi:10.1016/j.ecresq.2006.09.003

McClelland, M. & Cameron, C. (2012). Self-regulation in early childhood: Improving conceptual

clarity and developing ecologically valid measures. *Child Development Perspectives, 6,* 136-142. doi:10.1111/j.1750-8606.2011.00191.x

McClelland, M., Cameron, C., Connor, C., Farris, C., Jewkes, A. & Morrison, F. (2007). Links between behavioral regulation and preschoolers' literacy, vocabulary, and math skills. *Developmental Psychology, 43,* 947-959. doi: 10.1037/0012-1649.43.4.947

McClelland, M., Geldhof, J., Cameron, C., & Wanless, S. (2015). Development and Self-Regulation. In R. Learner (Ed.), *Handbook of Child Psychology and Developmental Science, 7th Edition* (1-43). John Wiley & Sons, Inc.

Moilanen, K., Shaw, D., Dishion, T., Gardner, F., & Wilson, M. (2009). Predictors of longitudinal growth in inhibitory control in early childhood. *Social Development, 19,* 326-347. doi: 10.1111/j.1467-9507.2009.00536.x.

National Education Goals Panel (1995). *Reconsidering children's early development and learning: toward common views and vocabulary*. Retrieved from: http://govinfo.library.unt.edu/negp/reports/child-ea.htm

Neuenschwander, R., Rothlisberger, M., Cimeli, P., & Roebers, C. (2012). How do different aspects of self-regulation predict successful adaptation to school? *Journal of Experimental Child Psychology, 113,* 353-371. doi: http://dx.doi.org/10.1016/j.jecp.2012.07.004

Oregon Department of Education. (2015). *Oregon Kindergarten Assessment, Report Overview*. Salem, Oregon: Public Service Building.

Pas, E. & Bradshaw, C. (2014). What affects teacher ratings of student behaviors? The potential influence of teachers' perceptions of the school environment and experiences. *Prevention Science, 15,* 940-950. doi: 10.1007/s11121-013-0432-4

Ponitz, C. 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.

Sabol, T. & Pianta, R. (2012). Patterns of school readiness forecast achievement and socioemotional development at the end of elementary school. *Child Development, 83,* 282-299. doi: 10.1111/j.1467-8624.2011.01678.x

Schmitt, S., Pratt, M., & McClelland, M. (2014). Examining the validity of behavioral self-regulation tools in predicting preschoolers' academic achievement. *Early Education and Development, 25,* 641-660. doi:10.1080/10409289.2014.850397

Sektnan, M., McClelland, M., Acock, A., & Morrison, F. (2010). Relations between early family risk, children's behavioral regulation, and academic achievement. *Early Childhood Research Quarterly, 25,* 464-479. doi: 10.1016/j.ecresq.2010.02.005.

Shields, K. A., Cook, K. D., & Greller, S. (2016). *How kindergarten entry assessments are used in public schools and how they correlate with spring assessments* (REL 2017–182). Washington, DC: U.S. Department of Education, Institute of Assistance, Regional Educational Laboratory Northeast & Islands. Retrieved from http://ies.ed.gov/ncee/edlabs.

Shoda, Y., Mischel, W., & Peake, P. (1990). Predicting adolescent cognitive self-regulatory competencies from preschool delay of gratification: Identifying diagnostic conditions. *Developmental Psychology, 26,* 978-986.

Smith-Donald, R., Raver, C. C., Hayes, T., & Richardson, B. (2007). Preliminary construct and concurrent validity of the Preschool Self-regulation Assessment (PSRA) for field-based research. *Early Childhood Research Quarterly*, *22*(2), 173-187.

U.S. Department of Education and Department of Health and Human Services (2013). *Race to the top- Early learning challenge (RTT-ELC)* (Catalog of Federal Domestic Assistance (CFDA) Number: 84.412A.). Washington, D.C.: U.S. Federal Register, Volume 78, No. 169.

U.S. Department of Health and Human Services (2010). *The head start child development and early learning framework* (Office of Head Start, Administration for Children and Families Contract no. HHSP233201000415G). Arlington, VA: Head Start Resource Center.

Wakschlag, L. S., Briggs-Gowan, M. J., Hill, C., Danis, B., Leventhal, B. L., Keenan, K., ... & Carter, A. S. (2008). Observational assessment of preschool disruptive behavior, part II: validity of the Disruptive Behavior Diagnostic Observation Schedule (DB-DOS). *Journal of the American Academy of Child & Adolescent Psychiatry*, *47*(6), 632-641.

Wanless, S., McClelland, M., Tominey, S., Acock, A. (2011). The influence of demographic risk factors on children's behavioral regulation in prekindergarten and kindergarten. *Early Education and Development, 22*, 461-488. doi: 10.1080/10409289.2011.536132

Wayman, J.C. & Jimerson, J.B. (2013). Teacher needs for data-related professional learning. *Studies in Educational Evaluation, 42,* 25-34. doi:10.1016/j.stueduc.2013.11.001

Weiland, C., Wolfe, C.B., Hurwitz, M.D., Clements, D.H., Sarama, J.H. & Yoshikawa, H. (2012). Early mathematics assessment: validation of the short form of a prekindergarten and kindergarten mathematics measure. *Educational Psychology, 32*(3), 311-333.

Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): A method of assessing executive function in children. *NATURE PROTOCOLS-ELECTRONIC EDITION-*, *1*(1), 297.

Zelazo, P. D., & Carlson, S. M. (2012). Hot and cool executive function in childhood and adolescence: Development and plasticity. *Child Development Perspectives*, *6*(4), 354-360.

Table 1
Descriptive Statistics for the Study Participants (N= 1864) *and Variables*

Variable		*Frequency* (%)	*Mean*	*SD*	*Range* *Min*	*Max*
Control variables						
Student age (months)			64.96	4.28	48	101
Ethnicity	African American	32.7				
	Caucasian	51.2				
	Hispanic	7.2				
	Native American	0.5				
	Asian	3.5				
	Multiple	4.2				
	Other	0.6				
Gender	Male	52.0				
	Female	48.0				
Early Language Learner (ELL)	No	91.4				
Predictor Variables: Students' Self-Regulation[1]						
Directly Assessed Frustration Tolerance			0.00	1.00	-3.04	1.34
Directly Assessed Delay of Gratification			0.00	1.00	-5.41	3.14
Teacher Report of Self-Regulation (CBRS)			0.04	1.00	-2.97	1.45
Outcome Variables: Academic Achievement[2]						
Mathematics Achievement (TEAM-SF)			100.72	14.58	48.89	141.08
Literacy Achievement (PALS)			100.31	15.01	60.73	124.09

[1] Self-Regulation measures are reported as z-scores (mean= 0, standard deviation= 1)

[2] Outcome variables are reported as standardized scores (mean= 100, standard deviation= 15)

Table 2
Correlation Matrix of Predictor and Outcome Variables

	1	2	3	4	5
1. Teacher-Reported Self-Regulation	--	**.241**	**.108**	**.431**	**.442**
2. Directly Assessed Frustation Tolerance		---	**.125**	**.369**	**.292**
3. Directly Assessed Delay of Gratification			---	**.138**	**.077**
4. Mathematics				---	**.561**
5. Literacy					---

Note. All correlations are bolded and significant at $p < .01$.

Table 3
Summary of Simple Regression Analysis for Variables Predicting Students' Academic Achievement

	Math Achievement *(TEAM-SF)*		**Literacy Achievement** *(PALS-K)*	
Fixed Effects				
	Est.	*SE*	Est.	*SE*
Intercept	85.16***	4.23	87.99***	4.57
Covariates				
Gender (boy)	2.26***	0.56	0.70	0.60
Early Language Learner Status	-2.19*	1.13	-3.62**	1.21
White	2.41***	0.60	-0.19	0.64
Age	0.19**	0.06	0.19**	0.07
Assessments of Self-Regulation				
Teacher Report (*CBRS*)	12.79***	1.82	9.21***	1.97
Frustration Tolerance (*SELweb*)	10.19***	1.82	4.43*	1.97
Assessments Interaction				
Teacher Report * Direct Assessment	-11.26***	2.86	-3.83	3.09

Note: All models accounted for nesting of children within classrooms. All estimates are standardized. *$p < .05$ **$p < .01$; ***$p < .001$.

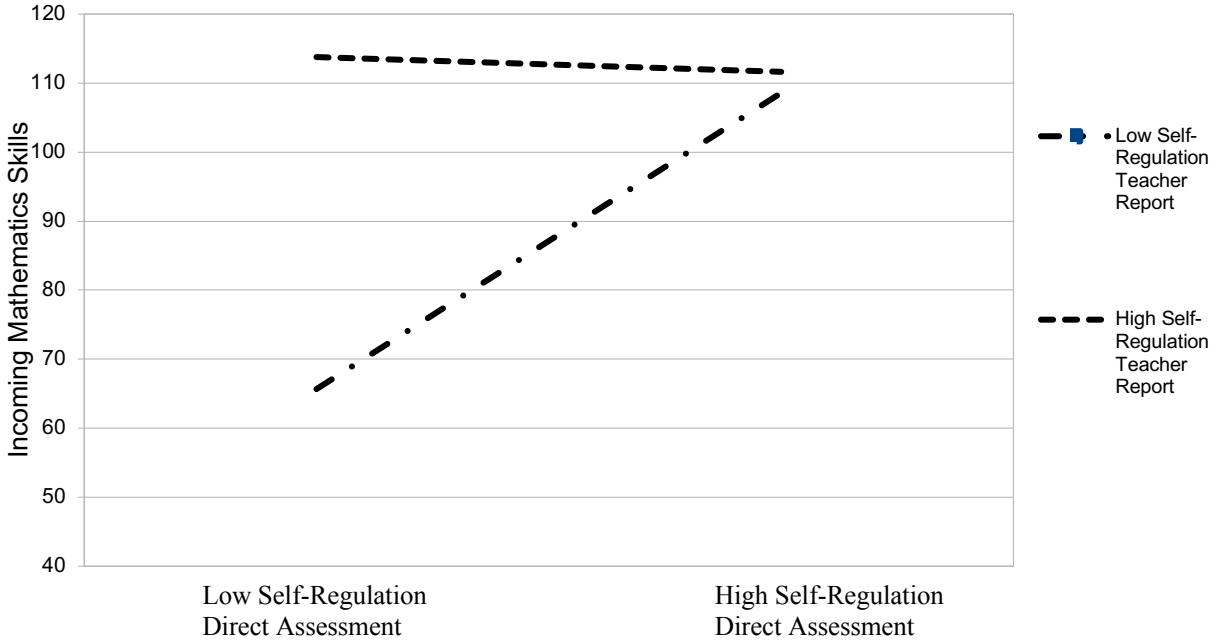

Figure 1. Students' predicted mathematics achievement at kindergarten entry. The effect of the interaction between students' directly assessed self-regulation and teacher reported self-regulation on students' predicted mathematics scores at kindergarten entry.

Examining the Validity of a Widely-Used School Readiness Assessment: Implications for Teachers and Early Childhood Programs

Jaclyn M. Russo [1]
Amanda P. Williford [1]
Anna J. Markowitz [2]
Virginia E. Vitiello [1]
Daphna Bassok [2]

[1]University of Virginia, Center for Advanced Study of Teaching and Learning (CASTL)
405 Emmet Street South, Ruffner Hall, Room 221
Charlottesville, VA 22904

[2]EdPolicy Works
PO Box 400879
Charlottesville, VA 22904
434-243-2021

(Published in *Early Childhood Education Quarterly,* 2019)

Russo, J. M., Williford, A. P., Markowitz, A. J., Vitiello, V. E., & Bassok, D. (2019). Examining the validity of a widely-used school readiness assessment: Implications for teachers and early childhood programs. *Early Childhood Research Quarterly*, *48*, 14-25. doi: 10.1016/j.ecresq.2019.02.003

Abstract

This study explored the validity of a widely-used, performance-based assessment of children's school readiness skills in the fall and spring of preschool. Using a sample of 1,109 children (mean age in the fall= 4.54 years; *SD*= 3.69 months) in 90 classrooms, we compared children's school readiness skills as assessed by teachers using Teaching Strategies GOLD (TS GOLD) to readiness skills as assessed by independent data collectors using standardized, direct assessments. Findings indicated evidence of convergent validity: TS GOLD scores were significantly associated with other assessments of similar skills. Evidence of discriminant validity was limited: TS GOLD domains were highly associated with one another and did not show differentiation in predicting direct assessment scores. In addition, comparison of intraclass correlations (ICCs) showed that children's skills were estimated as being much more similar to one another within a classroom when assessed using TS GOLD as compared to the direct assessments. More research is needed to ensure psychometrically sound readiness assessments, and prior to making strong policy and practice recommendations.

Keywords: *early childhood, school readiness, performance-based assessment, construct validity, skill differentiation*

Examining the Validity of a Widely-Used School Readiness Assessment: Implications for Teachers and Early Childhood Programs

Recent research highlighting the importance of children's skills at kindergarten entry for short and long-term well-being (McClelland, Acock, & Morrison, 2006; Sabol & Pianta, 2012) has spurred substantial interest among policymakers and practitioners in measuring and improving these skills. School readiness is a multidimensional construct, encompassing a set of interrelated skills across a range of early learning domains (e.g., academic, executive functioning, language, social-emotional, and behavioral; Boivin & Bierman, 2014; Bulotsky-Shearer & Fantuzzo, 2011; National Education Goals Panel, 1999; Sabol & Pianta, 2017; Schmitt, Pratt & McClelland, 2014). At kindergarten entry, a significant proportion of children lack school readiness competencies across foundational academic skills (Fitzpatrick, McKinnon, Blair & Willoughby, 2014; Janus & Duku, 2007), and these early deficits are predictive of children's later academic success in school. Children who enter kindergarten with fewer school readiness skills develop fewer skills over the course of elementary school compared to those who enter kindergarten solidly displaying these foundational skills (e.g., Duncan et al., 2007; Goldstein, McCoach, Yu, 2016; Hattie, 2008; Heckman & Mosso, 2014).

In response, there is now increased attention to measuring and monitoring young children's readiness skills during preschool and kindergarten. Indeed, several large-scale policy initiatives, such as the Race to the Top Early Learning Challenge and Quality Rating and Improvement Systems include specific provisions around school readiness assessments. Nationally, early childhood programs are using multiple methods to assess children's readiness skills, including performance-based, observational measures; rating scales; and direct assessments (Daily, Burkhauser & Halle, 2010; Friedman-Krauss et al., 2018). The most widely-used assessment tools in early childhood are performance-based, observational measures

(Friedman-Krauss et al., 2018), which require teachers to observe their students' skills across all learning domains during the school day. These assessments are often selected for use in early education settings because they 1) assess children's skills during regular, everyday activities, 2) are comprehensive in terms of measuring readiness skills across multiple learning domains, and 3) provide on-going, formative data to inform teachers' instructional practice over the school year (Riley-Ayers, 2014; Snow, 2006; Snow & Van Hemel, 2008).

Despite their popularity, there is limited research evidence as to whether or not performance-based measures achieve these aims. Some have asserted that performance-based measures may be difficult for teachers to administer, leaving scores susceptible to systematic differences in the way in which teachers assign scores to children (e.g., too lenient or too severe), and creating difficulty in discriminating across skills (Ackerman, 2018; Engelhard, 1994). Previous research examining the psychometric properties of performance-based, observational measures have yielded mixed findings (Lambert, Kim, & Burts, 2015; Miller-Bains, Russo, Williford, DeCoster & Cottone, 2017; Soderberg et al., 2013). When compared to direct assessments, these measures show evidence of convergent validity (i.e., skills assessed are associated with other assessments of theoretically similar skills; Lambert et al., 2015; Miller-Bains et al., 2017), however evidence of discriminant validity (i.e., weaker associations between theoretically dissimilar skills) is limited (Miller-Bains et al., 2017). Additionally, recent research examining the validity of performance-based measures has been cross-sectional, restricting our understanding of how these measures function over time—a critical aspect of measures that are intended to monitor the progress of growth of children's skills across the school year. In the present study, we examine how well a widely-used, performance-based readiness assessment measures and distinguishes between readiness skills from fall to spring across one preschool year

as compared to a battery of previously validated, widely-used direct assessments, allowing us to explore if the psychometric properties of the assessment are consistent over time, and how gains in children's readiness skills using teacher-rated performance-based assessments compares to gains in children's skills when independently assessed using well-established, standardized direct assessments.

School Readiness

School readiness is a term used to denote foundational skills across a range of early learning domains, including: cognitive (children's content knowledge); receptive language (listening and understanding others' language); expressive language (producing language that effectively communicates); executive functions (attention control and shifting, response inhibition, and working memory); and social-emotional and behavioral skills (cooperation, sharing, behavioral regulation; Blair & Raver, 2015; Boivin & Bierman, 2014; Kagan, Moore & Bredekamp, 1995; Regenstein, Connors, Romero-Jurado, & Weiner, 2017; Sabol & Pianta, 2017). Although distinct, these skills are interrelated, and studies have shown moderate correlations among academic and social-emotional domains (Bierman et al., 2008; McClelland et al., 2007). Further, children vary in their skills across domains and these skills work together in order to promote further academic skill development (Sabol & Pianta, 2012, 2017). For example, a child who is able to manage emotions and behaviors and interact successfully with peers is more able to participate in learning activities, and thus more able to develop early literacy and math skills compared to children without these skills (Clements, Sarama, & Germeroth, 2016; McClelland et al., 2007). Thus, school readiness can be conceptualized as a constellation of foundational skills that a child is able to use in combination to succeed in school.

Measuring School Readiness Skills in Early Childhood

Given the importance of early skills for later success, there has been increasing emphasis on using assessments to measure young children's school readiness skills both in preschool and at kindergarten entry. Indeed, all 50 states including the District of Columbia have developed Early Learning Guidelines (ELG) for early childhood education (ECE) programs to promote the development of foundational readiness skills in preschool (Daily et al., 2010; Ackerman & Coley, 2012). Further, over 70% of states implement some form of readiness assessment at the beginning of kindergarten (Ackerman & Coley, 2012; Connors-Tadros, 2014; Shields, Cook, & Greller, 2016).

Performance-based assessments in early childhood programs. Due to the proliferation of assessments available and broad set of assessment goals, the National Research Council (Snow & Van Hemel, 2008) published recommendations for school readiness assessments arguing that 1) readiness assessments should measure children's skills individually across foundational early learning domains (i.e., cognitive, language, executive functioning, social-emotional and behavioral) and, 2) that the intended use of the assessment data be clearly delineated for teachers, parents and programs (Regenstein et al., 2017; Snow & Van Hemel, 2008). For example, assessment data that are used for planning activities and tracking learning can be collected at the individual child level to identify areas of strength or need, or aggregated across children at the classroom level to monitor the effectiveness of an intervention program (Snow & Van Hemel, 2008). Programs are encouraged to use more than one assessment tool if there are multiple objectives (Regenstein et al., 2017).

Given the variety of goals for these assessments, there are multiple types of assessments used to measure children's readiness skills, including direct assessments, teacher rating scales, and performance-based, observational measures. The most commonly used assessment type in

early childhood settings are performance-based, observational measures, also known as authentic or work sampling tools (Ackerman & Coley, 2012; Friedman-Krauss, 2018). Performance-based, observational measures require teachers to collect data from several sources, including observational notes and samples of children's work. Teachers then complete checklists and scales to identify children's ability levels relative to developmental expectations (Heroman, Burts, Berke, & Bickart, 2010; Meisels, Bickel, Nicholson, Xue, & Atkins-Burnett, 2001; Riley-Ayers, 2014). Using this process, teachers obtain information across multiple early learning domains (Heroman et al., 2010). Performance-based assessments are intended to be used repeatedly, such that over time teachers have a broad, ecologically valid view of children's development (Riley-Ayers, 2014; Snow, 2006; Snow & Van Hemel, 2008). This type of assessment is referred to as formative because the data are intended to be used by teachers to support children's learning needs through targeted instruction and close monitoring of progress across learning goals (Riley-Ayers, 2014).

Evidence for validity of performance-based assessments. Validity is a fundamental component of test construction and refers to the "degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests" (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999, p. 9). Despite their popularity, recent research has raised concerns that children's scores on performance-based, observational measures may be more attributable to systematic differences in teachers' ratings than to differences in children's skills (Engelhard, 1994; Waterman, McDermott, Fantuzzo, & Gadsden, 2012). Waterman et al. (2012) found that 27.6% of the variation from a child observation tool was unrelated to actual child differences. Raters may assign scores that are too lenient or too severe or be less able to

discriminate across skills (i.e., accurately rating a child's math versus literacy skills; Engelhard, 1994; Miller-Bains et al., 2017). If this is the case, then the data are less useful in fulfilling their intended purposes—for example guiding teacher instruction, modifying teacher professional development, adjusting curricula, and tracking program level outcomes (for a comprehensive list of how programs are using performance-based assessment data see Schilder & Carolan, 2014).

As use of performance-based, observational measures grows, it becomes increasingly important to understand the key measurement properties of these tools. They must be able to both measure the skills that they are claiming to measure (i.e., convergent validity) and distinguish between different types of readiness skills (i.e., not measure skills they are *not* intended to measure, known as discriminant validity)—two important aspects of construct validity. In addition to displaying validity at any given time point, these measures also need to measure changes in children's skills over time.

Currently, the most widely-used performance-based, observational measure in ECE programs is Teaching Strategies GOLD (TS GOLD, Heroman et al., 2010). Nine state preschool programs require or encourage the use of TS GOLD (Schilder & Carolan, 2014), four states are currently piloting the use of TS GOLD as part of their Kindergarten Entry Assessment (KEA) systems, and five states are in other phases of implementation of the measure (Weisenfeld, 2017). Results from a telephone survey of 73 Head Start and Early Head Start program directors revealed that 60% of programs were using some form of the TS GOLD assessment system (Isaacs et al., 2015). TS GOLD is designed to help teachers create a developmental profile for each child in order to scaffold his or her learning.

Several studies have explored the psychometric properties of TS GOLD in preschool samples (e.g., Burts & Kim, 2014; Lambert et al., 2015; Lambert, Kim, & Burts, 2014). Lambert

and colleagues examined the convergent validity of TS GOLD in a large sample of 3-and 4-year-olds by estimating correlations between TS GOLD domains and other well-established direct assessments of similar constructs at the same time point. Correlations between TS GOLD and the direct assessments were found to be positive and moderate within readiness domains (i.e., TS GOLD math score was significantly and positively associated with a direct assessment of math skills). However, they also noted significant associations across domains (i.e., TS GOLD literacy was highly associated with the direct assessment of math skills).

Lambert et al. (2015) further probed the high level of cross-domain correlation by comparing the intra class correlations (ICCs)—which measure the amount of variability in children's scores attributable to the classroom level—associated with TS GOLD domains and found large ICCs associated with TS GOLD (ranging from .38 to .54), indicating that between 38 to 54% of the variance between children in the sample was attributable to what classroom the children resided in at the beginning of the year. In other words, children looked more similar within a classroom than would be expected at the beginning of the school year given that children have diverse experiences prior to formal schooling; suggesting teachers' scores of children's readiness skills using TS GOLD may reflect information (including teacher bias) that is independent of children's skill levels. In comparison, ICCs for direct assessments (e.g., early childhood math and literacy assessments) are often much smaller (i.e., approximately .10- .20; Schmitt et al., 2014). Similarly, an observation of children's time spent engaged in learning revealed an ICC of .14 (Rimm-Kaufman, Curby, Grimm, Nathanson, & Brock, 2009).

The validity of TS GOLD, including examination of the ICCs, was recently explored in a large sample of kindergarten students. Results of this study found that while corresponding domains (i.e., TS GOLD literacy and direct assessment of literacy skills) were concurrently and

significantly associated with one another, there was a lack of evidence for discriminant validity among the TS GOLD domains (i.e., high associations between all domains). Moreover, the ICCs for TS GOLD were large (ranging from .19 to .59), whereas the ICCs for the direct assessments were much smaller (ranging from .02 to .04; Miller-Bains et al., 2017). The lack of differentiation across early learning domains (i.e., a child's skill level looks very similar across literacy, math, language, social) and between children within a classroom (children's scores are similar if they share the same class) of TS GOLD when compared to direct assessments calls into question the utility of the measure to be used to accurately assess children's skills comprehensively across learning domains.

Notably, however, these validation studies are all limited to measurement at a single time point at the beginning of either the preschool or kindergarten year (Burts & Kim, 2014; Miller-Bains et al., 2017). One of the purported strengths of TS GOLD and other performance-based measures is that they are designed such that teachers could assess children's skills across a school year. TS GOLD has assessment windows at 3-points across a school year in order to track children's growth comprehensively across learning domains. However, to date, we know of no research that has examined the validity of TS GOLD longitudinally. Additionally, research has yet to explore how well children's school readiness gains as measured by TS GOLD are associated with gains in children's skills as measured by direct assessments. Understanding how performance-based, observational measures perform across time is critical given that they were designed to provide teachers with on-going data in order to inform instructional practices, to help teachers to understand how students are growing across skills, and for program-wide progress monitoring.

Present Study

This study fills an important gap in our understanding of the validity of one of the most widely-used, performance-based, observational readiness assessments in ECE settings. By probing the convergent validity, discriminant validity, and the classroom-level variance of TS GOLD in the fall, spring, and over the course of the preschool year, this study will inform policymakers and practitioners hoping to use TS GOLD to accurately capture children's growth in readiness skills across multiple learning domains.

First, we explored the convergent validity of TS GOLD relative to a set of well-validated direct assessments of children's school readiness skills. These comparisons are conducted at fall and spring to provide insight into how teachers' use of TS GOLD changes over time. Second, we explored the discriminant validity of TS GOLD, both in terms of distinguishing between individual school readiness skills and among the skills of children within a classroom. Third, we examined the ICCs to explore the amount of variance in both the TS GOLD and direct assessment scores that is attributable to a child's classroom membership. Finally, we examined both concurrent and discriminant validity when comparing students' skill gains on direct assessments with students' skill gains as measured by TS GOLD, a question which has not been examined in previous research. Given prior research, we hypothesized that TS GOLD would demonstrate evidence of convergent validity but demonstrate limited evidence of discriminant validity with direct assessments at fall and spring and also when examining gains from fall to spring (Miller-Bains et al., 2017). Descriptively, we were interested in examining whether evidence of discriminant validity would be stronger at spring and when examining gains over the school year.

Method

Participants

The sample for this study comes from data collected as part of the pilot of a large, statewide Quality Rating and Improvement System (QRIS) in Louisiana during the 2014-2015 school year. As part of Louisiana's 2012 Education Reforms, leaders in the state passed the Early Childhood Education Act, which was an effort to create a cohesive early childhood education system focused on providing high-quality learning opportunities and improving school readiness statewide.

As part of the study of the Early Childhood Education Act, a total of 1,109 children were recruited from 90 preschool programs (1 classroom per program). Of the 90 preschool programs, 56.7% were state-funded preschool programs, 20.0% were federally-funded Head Start programs, 12.2% were private, non-religious early childhood programs and 11.1% were private, religious-based preschool programs. The sample used in the present study included all available data from all recruited children. On average, children were from low-income families with relatively low levels of parent education: 71.5% of children were Black, and 10.4% of children came from families that spoke a second language at home (complete sample characteristics are provided in Table 1).

Recruitment and attrition. The university IRB approved all study procedures. The research team collaborated with the state Department of Education to select five Louisiana parishes participating in the pilot phase of the reform that captured the geographic and demographic diversity of the state. From a list of all preschool programs receiving public funds across the five parishes, the research team randomly selected 90 programs, stratified by parish and program type. Within each program, all teachers of classrooms serving primarily four-year-olds and typically-developing children were randomly ordered and the first teacher from each program was contacted. Six teachers declined to participate or were later found to be working in

classrooms that did not serve primarily four-year-olds, so the teacher in the next eligible classroom on the randomized list was contacted. If there were no other eligible teachers at the program, the program was dropped and the next randomly-selected program within that parish/program type was contacted as a replacement.

All teachers in the sample were female. Fifty-four percent were White, 38% were Black, 2% were Hispanic, and 6% reported another race/ethnicity. The majority of teachers in the sample had at least a 4-year degree; 65% of the sample had attained a BA and 28% of the sample had attained more than a BA. Teachers had, on average, 15 years of experience, but there was substantial variability around this mean (range=0-39, *SD*= 11).

All parents in the selected study classrooms received a letter informing them of the study, asking them to complete a parent survey, and providing the option to opt their child out of assessment by contacting their child's teacher, their program director, or the research team. Approximately 80% of parents responded to the parent survey in the fall. Four weeks into the school year, 6 boys and 6 girls were randomly selected from the pool of families whose parents returned the parent survey per classroom—all children were required to be 4-years-old on or before September 30, 2014, not have an active IEP, and had receptive and expressive skills in English sufficient to participate in the assessment (based upon teacher report). In the spring, data collectors were provided with a roster of the names of the children who were assessed in the fall and remaining children whose parents completed a survey were listed in random order at the bottom of the list. If a child needed to be replaced in the spring (no longer in the classroom, refused to participate, absent for more than 3 months during the school year or absent during our efforts to collect data), the data collectors used the randomized list of remaining children to select a replacement child—90% of the total recruited sample received some form of a direct

assessment in the spring. Missingness was not related to any child or family characteristics, including age, gender, race, first language, family income, parent education, single parent status, or household size.

Data Collection Procedures

Teachers assessed children's skills using TS GOLD as part of their normal educational practice, independently from the research team, in order to better understand children's growth in school readiness skills across the preschool year and to individualize their instruction to support student learning. As part of their participation in the pilot, teachers were provided training on TS GOLD through the Department of Education and their local early childhood network. Lead teachers completed TS GOLD in the fall and spring. Teachers were required to assess all children who were receiving care supported by public funds as part of the state's QRIS; thus, all children in Head Start and state preschool classrooms were assessed, and all children whose families paid for care in part with child care subsidy dollars were assessed in private center-based care. In all but four cases, lead teachers were the same in the fall and spring; the four teachers who left their classrooms during the year were replaced, and the new teacher completed TS GOLD in the spring.

The research team hired data collectors to conduct the direct assessments on randomly selected children within participating classrooms. Data collectors local to Louisiana were recruited (through online job postings), hired, trained, and supervised by the research team. Data collectors participated in a week-long training for the larger study with two days focused exclusively on the direct assessment measures. This training focused both on the procedures for standard administration as well as guidelines for testing young children (e.g., how to ask young children for verbal assent, how to gauge attention to determine if a child needs a break, and when

to discontinue testing). All data collectors passed certification tests on all instruments prior to data collection. In order to be certified, each data collector had to administer and film the assessment battery with a four-year-old child. The videos were then reviewed by the research team to ensure data collectors were able to administer the assessments according to the standardized protocol. Additional training and calibration were provided throughout the data collection windows as needed and data collectors conducted a short "refresher training" before the spring assessments.

Data collectors individually assessed children's language, literacy, mathematics and executive functions during the school day during the fall (September- November) and spring (March-June) in a single session lasting approximately 45 minutes (if needed the assessment battery could be split into multiple sessions) in a quiet location away from classroom distractions. As part of the testing protocol, data collectors described the direct assessment tasks in child friendly language and informed children that they could ask for a break or to stop testing at any time and data collectors were trained to take a break from or discontinue testing as needed based on the child's attention and comfort. Measures are described in more detail below, and descriptive information is provided in Table 2.

Measures

Teaching Strategies (TS) GOLD. TS GOLD is a multidimensional, performance-based, observational assessment where teachers are required to observe children's skills during typical instruction across nine broad areas of development (i.e., literacy, mathematics, language, social-emotional, cognitive, physical, science and technology, social studies and arts; Heroman et al., 2010). Teachers provide documentation of children's skills across these areas using their observations and evidence from the classroom such as notes or samples of student work. Using

an online portal, teachers rate children on each area of development using "age-bands" at several time points during the school year. To assist teachers in their rating of children's skills, TS GOLD provides 38 objectives across the nine skills and 65 behavioral indicators, referred to as "dimensions." Dimensions are organized on a 10- point scale with anchors at points 2, 4, 6 and 8 to reflect designated developmental expectations in each learning objective based on nationally normed benchmarks. Level 9 represents "beyond kindergarten expectations" (Heroman et al., 2010). See Data Appendix B for detailed information regarding the scoring of the TS GOLD domains.

In the present study, children's scores on the first five areas of development (literacy, mathematics, language, social-emotional, and cognitive) were used. Scores were standardized to have a mean of zero and a standard deviation of one to create a common metric for comparison with the direct assessments. Previous studies of TS GOLD have demonstrated evidence of moderate to strong internal reliability (Lambert et al., 2015). TS GOLD data were reported by teachers in their final form to the state, and the state provided us with these finalized scores. We were unable to examine Cronbach's alpha for TS GOLD in our sample because we did not have access to the data underlying TS GOLD scores.

Direct Assessments of School Readiness Skills. A more in-depth description of the skills assessed across the direct assessments as well as the administration details can be found in Appendices A and C, respectively.

Language. To measure children's early receptive language skills, two assessments were used, the Peabody Picture Vocabulary Test, 4th Edition (PPVT-IV; Dunn & Dunn, 2007) and the Picture Vocabulary subtest of the Woodcock-Johnson III Tests of Achievement (WJ-III;

Woodcock, McGrew, & Mather, 2001). These scales were moderately correlated in both the fall (r= .57) and spring (r= .65).

Children were assessed using the 4th edition of the PPVT (Dunn & Dunn, 2007). Children were shown four pictures and asked to identify the picture corresponding to a word verbally stated by the administrator. The PPVT-IV demonstrates evidence of strong reliability in previous research (Cronbach's alpha = .94, Dunn & Dunn, 2007), and in the present sample (Cronbach's alpha for the fall administration= .91, in the spring= .92).

The Picture Vocabulary subtest of the Woodcock-Johnson III Tests of Achievement (WJ-III; Woodcock et al., 2001) was administered to assess children's expressive language abilities. In this Picture Vocabulary test, children were required to verbally name visual pictures presented to them. This test has demonstrated strong reliability both in previous research (Cronbach's alpha > .80; Woodcock et al., 2001) and in the current sample (alpha= .78 in the fall, and .77 in the spring).

Literacy. Children's emergent literacy skills were assessed using two subtests from the Test of Preschool Early Literacy (TOPEL; Lonigan, Wagner, Torgesen, & Rashotte, 2007): Phonological Awareness, which assesses children's ability to manipulate individual sounds (phonemes) and Print Knowledge, which assesses children's knowledge of the alphabet, written language conventions, and writing form. The TOPEL shows evidence of strong reliability (Cronbach's alpha > .80) and criterion validity in both previous research (Lonigan et al., 2007) and in the present sample (Cronbach's alpha for Phonological Awareness was .74 and .78 in the fall and spring, respectively; for Print Knowledge it was .88 and .87, respectively). These scales were moderately correlated, r= .41 in the fall and r= .51 in the spring.

Math. Two subtests from the WJ III-TA were administered to measure children's early mathematics skills: Applied Problems, which measures children's ability to analyze and solve math problems, and Quantitative Concepts which assesses a child's knowledge of basic math concepts and number patterns. As mentioned above, the WJ III-TA shows evidence of strong reliability and validity (Woodcock et al., 2001), and this was true in this sample as well (Cronbach's alpha for Applied Problems was .83 and .84 in the fall and spring, respectively; for Quantitative Concepts it was .87 and .89, respectively). The correlation for these tests was somewhat higher than for other domains, but still modest, r= .60 in the fall and r= .66 in the spring.

Executive Functions. Children completed two widely-used tasks to directly assess executive functions: the Pencil Tap test and Head Toes Knees Shoulders (HTKS). The Pencil Tap measures inhibitory control and requires children to tap a pencil once when the examiner tapped twice and vice versa (Smith-Donald, Raver, Hayes, & Richardson, 2007). The Pencil Tap demonstrates evidence of concurrent validity with other established measures of inhibitory control and strong inter-rater reliability (Rimm-Kaufman et al., 2009; Smith-Donald et al., 2007). HTKS measures inhibitory control, working memory, and attention by asking the child to do the opposite of what the examiner says (e.g., touch your head when I say "Touch your toes," Ponitz, McClelland, Matthews, & Morrison, 2009). HTKS shows evidence of concurrent and construct validity and inter-rater reliability (Ponitz et al., 2008). In the fall, these scales were correlated r= .45, in the spring r= .43. As with most direct assessments, the direct assessments of executive functions measure a set of skills that are narrower in breadth compared to the TS GOLD cognitive and social-emotional domains. However, the underlying skills—for example, inhibition of an immediate response during HTKS and being able to wait one's turn during circle time—are

aligned across the direct assessments and TS GOLD. Further, these direct assessments have been used in previous validation research with performance-based, observational measures (Miller-Bains et al., 2017).

Analytic Strategy

Missing Data

Of the full sample of 1,109 children, not all had full data on all TS GOLD domains and all direct assessments. More specifically, 72% of children had complete data on the direct assessments, 69% of children had complete data on TS GOLD, and 53% had complete data for both direct assessments and TS GOLD. To address missingness we first compared the covariate information from the 1,109 children to those children with complete TS GOLD and direct assessment data (N=584). T-tests across all child- and family-level demographic characteristics, including age, race, gender, family income, parent education, single parent family, household size, and English as a second language, indicated no significant differences (all t-test p values > .05) between children missing data and those without missingness. However, there were some differences at the teacher level. With regards to TS GOLD, older teachers were less likely to complete assessment data, white teachers were marginally more likely to have complete assessment data (p<0.10), and other race teachers were less likely to have complete assessment data. Because of these scattered teacher-level differences, we use two missing data strategies.

First, to maximize all available data from the sample, all descriptive analyses (including means, standard deviations, correlations, and intraclass correlations) were calculated using the largest possible sample (sample sizes range from 768-1016). These largest possible samples are obtained through listwise deletion; variability in the sample sizes for all analyses are reported accordingly in the corresponding tables (Tables 1-4). Second, for our examination of the

relationships between student growth on the direct assessments and on the TS GOLD domains we used multiple imputation in order to be able to include the full sample of 1,109 tested children (Table 5). To impute data, we used multiple imputation by chained equations, and created 15 additional datasets. Because our analysis is not interested in the impact of an independent "treatment" variable on a dependent "outcome" variable, we imputed all missing data including both TS GOLD and direct assessment scores, an approach consistent with recent research on multiple imputation (e.g., Johnson & Young, 2011; Young & Johnson 2010; Zaninotto & Sacker, 2017). Notably, however, results were not sensitive to the use of multiple imputation. That is, in a follow-up analysis (not shown) in which we used the most restricted set of data (children had to have all assessment data, *N*=584) the pattern of results was the same for each analysis.

Creation of Direct Assessment Composite Scores

The direct assessments are scored on different scales with different ranges, means, and standard deviations, and so the scores for each are not directly comparable. Thus, to create the composite scores for language, literacy, mathematics and executive functions, we first took the raw scores from each test within a domain (e.g., the PPVT-IV and the WJ III-TA within the language domain) at each time point (e.g., fall and spring) and converted each into a z-score with a mean of zero and standard deviation of one. This conversion allows each test within a domain to contribute equally to the final composite score. We then took an average of the z-scores by time point, and then finally re-standardized (e.g., mean of zero, standard deviation of one). This final z-score allows us to interpret coefficients as effect sizes. This process was replicated for each domain.

Primary Analyses

The present study explored the convergent and discriminant validity of TS GOLD by probing cross-sectional associations between TS GOLD and direct assessments in the fall and spring, and associations between gains on each type of assessment. Gain scores on both the direct assessments and the TS GOLD domains were calculated by regressing children's spring scores on their fall scores; the gain score is the standardized residual from this regression. To assess convergent validity, we examined the correlation coefficients between the TS GOLD domains and the direct assessments within the same skill area (i.e., correlation between TS GOLD math and direct assessments of math) in the fall and spring, and for children's gains from fall to spring (Table 3). We expected correlations between assessments of the same skill area to demonstrate moderate to strong positive associations. In Data Appendix A, we provide detailed descriptions of the skills assessed across each TS GOLD domain and the direct assessments.

Next, to explore TS GOLD's ability to discriminate amongst different early learning domains we examined the correlations between the TS GOLD domains and the direct assessments of *different* learning domains (e.g., TS GOLD literacy vs. direct assessment math) in the fall, spring, and gain scores. Evidence for discriminant validity would be established if the association between different learning domains was weaker compared to the association between domains of same or similar constructs (Downing, 2003).

We also calculated the proportion of variance in TS GOLD and direct assessment scores that is attributable to variability between teachers by computing the intraclass correlations (ICCs) for both TS GOLD domains and the direct assessments. The ICC partitions the amount of variance that is between the clusters (e.g., attributable to being in a particular classroom) from the portion of the variance that is within the cluster (e.g., attributable to differences between the students within a classroom). The ICC is defined by the following ratio:

$$\frac{\sigma_B^2}{\sigma_B^2 + \sigma_W^2}$$

where σ_B^2 represents the between- cluster (i.e., between classrooms) variance and σ_W^2 represents the within-cluster (i.e., between students within the same classroom) variance. Descriptively, we compared the ICCs of TS GOLD subscales and the direct assessment composites. ICC coefficients were calculated in Stata (version 14).

Finally, to examine TS GOLD's discriminant validity in terms of gains over the preschool year across readiness domains, we performed a series of two-level models predicting spring direct assessment scores (controlling for fall direct assessment score) from each individual TS GOLD gain score. That is, we predict each spring direct assessment score (language, literacy, math, and executive function), from each TS GOLD domain gain score (language, literacy, math, cognitive, and socio-emotional) for a total of 20 individual regressions. As described above, these analyses were conducted on multiply imputed data using Stata 14.

Additionally, because of previous work demonstrating strong correlations among TS GOLD domains (e.g., Miller-Bains et al., 2017), we created a composite TS GOLD growth score by taking the mean of all standardized residuals across the TS GOLD domains. Then in individual models, we use this composite score to predict each individual spring direct assessment score; that is, in an additional 4 models, we regressed each spring direct assessment score on the composite TS GOLD score. In all models, we controlled for child level covariates, including age, sex, and ethnicity, as well as fall direct assessment scores (included as fixed effects). In addition, a random intercept was estimated for each classroom resulting in the estimation of separate variance terms at the classroom and child levels to account for nesting of children in classrooms in each model. For example, to examine children's gains in literacy we

entered children's spring directly assessed literacy scores as the outcome and fall direct assessment literacy score and gains on the TS GOLD literacy domain as the fixed effects, and classroom as the random intercept. This procedure was repeated for all learning domains, resulting in 24 independent mixed effects regression models (see Table 5). The equation for the mixed effects regression models is provided below:

$$Y_{ij}\ (\text{T2 DA score}) = \beta_{0j} + \beta_{1ij}(\text{T1 DA score}) + \beta_{2ij}\ (\text{TS GOLD gain score}) + r_{ij}$$

where Y_{ij} represents the spring direct assessment score for student *i* in classroom *j*; β_{0j}, represents the random intercept for classroom *j*; β_{1ij}, represents the fall direct assessment score for student *i* in classroom *j*; β_{2ij}, represents the TS GOLD gain score (standardized residual of Time 2 regressed on Time 1) for student *i* in classroom *j*; and r_{ij}, refers to child-level error. For ease of readability, we did not include covariates in the equation above; however, all models presented in Table 5 included the covariates listed above.

Results

Construct Validity

Convergent Validity. To establish the convergent validity of TS GOLD, we examined the correlations between each TS GOLD domain and the direct assessments in the fall and spring and over the preschool year (residualized gains). All correlation coefficients are presented in Table 3. Across skill areas, the within-domain correlations were positive and varied in strength by skill area ranging from modest to moderate. For example, the strongest positive within-domain correlations were observed for literacy, where associations between TS GOLD and the direct assessments were .42 and .53 in the fall and spring, respectively. Conversely, the associations between TS GOLD and the direct assessments were weakest within the cognitive and social-emotional domains at both time points and over the course of the year (.06- .23).

In comparing the fall and spring within-domain associations, associations were often stronger in the spring. For example, the association between children's language skills as assessed by TS GOLD and the direct assessments was .22 in the fall and .37 in the spring (*p* values < .001). Similar patterns were noted for math where the associations between TS GOLD and the direct assessments were .38 in the fall and .44 in the spring (*p* values < .001).

Within-domain correlations were smaller when examining children's skill gains than for either the fall or spring associations across all early learning domains. Similar to the fall and spring associations, the associations between TS GOLD and the direct assessments were the strongest for gains in literacy ($r = .20$). The weakest associations for gains were observed within the cognitive domain, and were non-significant ($r = .06$).

Discriminant Validity. To explore discriminant validity, we first compared cross-domain correlations to within-domain correlations. The cross-domain correlations were largely of the same magnitude as the within-domain associations in both the fall and spring, and for gains. For example, the magnitude of the relationship between the TS GOLD math domain and the direct assessments of language, literacy and math were very similar in the fall ($r = .27$, $r = .32$, and $r = .38$, respectively) and identical in the spring ($r = .44$, $r = .44$, and $r = .44$).

The cross-domain associations were consistently stronger in the spring as compared to the fall, as was observed for within-domain associations; that is, cross-domain associations did not diminish later in the academic year. For example, TS GOLD math was more strongly associated with direct assessments of literacy in the spring versus the fall ($r= .44$ and $r= .32$, respectively). That both within- and cross-domain associations were similar in magnitude at both fall and spring time points suggests that discriminant validity did not improve over the course of the school year.

ICC comparisons

To explore how well TS GOLD discriminates between students *within* a classroom as compared to the direct assessments, we compared the proportion of variance attributable to classroom membership versus the child for all TS GOLD domains to all direct assessments. To do this, we calculated the ICC coefficients of each assessment in the fall, spring, and also for gains, which are presented in Table 4. For the TS GOLD domains, the ICC or the proportion of variance attributable to the classroom ranged from .50 (literacy) to .62 (language and social-emotional) in the fall. In the spring, the ICC for TS GOLD remained large, ranging from .48 (literacy) to .63 (cognitive). These values indicate that 48% or more of a child's TS GOLD score are explained by their classroom membership. The magnitude of the ICCs for TS GOLD were similar in the fall and spring. The proportions of classroom variance for children's gains on TS GOLD ranged from .62 (language) to .71 (cognitive). These values indicate that at least 62% of the variance in children's skill gains are explained by children's classroom membership.

The proportions of classroom-level variance for the direct assessments were substantially smaller than those of TS GOLD. In the fall, these values ranged from .10 (math and executive functions) to .12 (literacy); in the spring, proportions of classroom variance ranged from .12 (language) to .21 (literacy). Overall, the proportion of classroom variance accounted for in the spring was slightly larger than in fall across all direct assessment domains. With regards to skill gains, there was wider variability, ranging from .09 (language) to .36 (literacy). Similar to TS GOLD, these values tended to be higher (with the exception of language), when compared to the fall or spring direct assessment scores.

Comparing Gains on TS GOLD to Gains on the Direct Assessments

As a final examination of how well gains on TS GOLD were associated with gains on

children's directly assessed skills, gains as measured by TS GOLD were used to predict children's spring direct assessment scores while controlling for children's fall direct assessment scores. We expected these results to mirror the skill gains correlation results. Importantly, however, these models are a more precise examination of how gains on TS GOLD are associated with *gains* on the direct assessments as we were able to account for nesting of children in classrooms and include key covariates. Results from these models are presented in Table 5. Recall that in these analyses each individual direct assessment score was regressed on each TS GOLD domain and the composite TS GOLD score *separately*. Thus, each coefficient in Table 5 represents the results of a different regression. Because all scores have been standardized, all coefficients can be interpreted as effect sizes.

With regard to convergent validity, these models indicated that gains as measured by TS GOLD across learning domains were associated with children's spring direct assessment scores. For example, gains in literacy as captured by TS GOLD was significantly associated with children's directly assessed literacy skills in the spring (b= .24, p< .001). However, as with the cross-sectional correlations, there was limited evidence of discriminant validity—meaning that within-domain predictions were *stronger* than cross-domain predictions, except for the literacy and social-emotional domains. Gains on TS GOLD in any domain was associated with all spring direct assessments, despite domain. For example, children's gains on TS GOLD math significantly predicted spring directly assessed language, literacy, math and executive functions. In a few cases the cross-domain predictions were larger or of identical magnitude to those of the within-domain (e.g., b = .18 for TS GOLD literacy to directly assessed math and b = .17 for TS GOLD math to directly assessed math). Moreover, predictions from the composite score (i.e., the average of all five TS GOLD gain scores) were large and statistically significant in all models.

Discussion

The present study used data from a statewide initiative to evaluate a new QRIS in a southern state in the U.S. The QRIS included multiple components, including the implementation of TS GOLD. The present study replicated and extended previous work (Lambert et al., 2015; Miller-Bains et al., 2017) examining the validity of a widely-used, performance-based, observational assessment, TS GOLD. We explored the convergent and discriminant validity of teachers' assessments of children's readiness skills using TS GOLD relative to independent data collectors' assessments of children's readiness skills using well-validated direct assessments in the fall and spring of the preschool year. Next, using ICC comparisons, we explored how well TS GOLD discriminates between students within a classroom compared to the direct assessments. Finally, we compared children's gains on direct assessments with children's gains as measured by TS GOLD, again examining evidence for convergent and discriminant validity. This study fills an important gap by increasing our understanding of the validity over time of one of the most widely-used, performance-based, observational readiness assessments in ECE settings. The sample in this study was comprised of a diverse group of children from low-income backgrounds who are often underrepresented in contemporary research. Thus, this study provides information about how school readiness skills are measured in a sample that is reflective of a population of children who are at risk for entering kindergarten without the school readiness skills needed for early and later school success (Garcia & Weiss, 2015).

Construct Validity

Convergent Validity. As expected and consistent with prior research, TS GOLD showed adequate convergent validity relative to direct assessments, in the fall, spring, and over the

course of the preschool year (e.g., Miller-Bains et al., 2017)—associations between children's readiness skills as measured by TS GOLD and the direct assessments within a skill area (e.g., literacy) were positive and modest to moderate in strength. As expected, we did find that associations between children's skills as measured by TS GOLD and direct assessments within a skill area were larger in the spring than in the fall, suggesting that over the course of the school year teachers' ratings of children using TS GOLD became more similar to students' directly assessed skills. However, in almost all cases the associations between TS GOLD and the direct assessments remained less than .50, indicating that there is a considerable amount of variability in TS GOLD scores that does not align with the directly assessed score for that skill. In previous research, associations between direct assessments of similar skills are strong (*r* values= .68 and .77 between direct assessments of early math skills; Willoughby, Blair, Wirth, & Greenberg, 2012). These results however, are consistent with prior research that has shown variability in the strength of associations between teacher report and direct assessments—some studies show associations as low as $r = .15$ (i.e., assessments of self-regulation; Ponitz; 2008), while others show associations as strong as $r = .76$ (i.e., assessments of early reading skills; Begeny, Eckert, Montarello, & Storie, 2008). One likely reason is the alignment between the items within each of the assessments. Most often, not only does the method differ (teacher report vs. direct assessment) but the content is also different in terms of the items. In this study, the direct assessments often contain items that assess more discrete skills (e.g., "name this letter") whereas in the teacher report the item may be more global (e.g., "child can identify capital letters"). Alternatively, teachers may have more pedagogical content knowledge or more direct interactions with children in certain content areas, making them better able to assess particular skills. For example, teachers receive substantial professional development related to the

development of early literacy skills and spend a lot of classroom time focused on early literacy (Early et al., 2010; Lobman, Ryan, & McLaughlin, 2005; Zaslow, Tout, Halle, Whittaker, & Lavelle, 2010) so it is perhaps not surprising that children's literacy skills as measured by TS GOLD and the direct assessments showed the strongest associations.

Discriminant Validity. We found that the magnitude of the associations among theoretically different domains (i.e., TS GOLD math and the direct assessments of literacy) was comparable to that of the magnitude of the within-domain associations (e.g., TS GOLD literacy with the direct assessments of literacy) indicating limited evidence for discriminant validity. For example, the TS GOLD literacy domain was equally and significantly associated with the direct assessments of literacy and math (i.e., both correlations were .42).

School readiness skills are inter-related, and prior research shows that within children skills across domains are positively correlated with the strength of these correlations ranging from modest to moderate (e.g., .69 for math and reading and .30 for attention and academic achievement; Duncan & Magnuson, 2011). For example, with regards to math, a student with strong self-regulation skills will be better equipped to inhibit an immediate response and think flexibly about many different solutions—demonstrating the interrelatedness of skills (Clements et al., 2016). However, while inter-related, previous research suggests that even in early childhood, direct assessments of children's learning skills show differentiation (e.g., Howes et al., 2007; Janus & Offord, 2007; Miller-Bains et al., 2017). If policy makers and program leaders expect teachers to capture across skill differences as part of their assessment practices, it is important that their assessment tools have the psychometric properties that will allow them to accurately capture within child variability across learning domains. Otherwise, these assessments require time that does not yield practical information for classroom instructional purposes.

Examination of the ICCs

We also compared ICCs of TS GOLD and the direct assessments to understand the extent to which variability in children's scores was explained by the nested structure of the data (here children within classrooms). These analyses showed that proportions of classroom-level variance for the TS GOLD domains were markedly larger than those of the direct assessments in the fall, spring, and across the preschool year. Approximately 50-60% of the variance in TS GOLD was attributable to classroom membership in both the fall and spring as compared to 10-20% across direct assessments in the fall and spring. Similarly, for children's gains, the TS GOLD estimates indicated that 48-63% of the variance was attributable to children's classroom membership—and again, the largest proportion of classroom-level variance for children's gains as measured by the direct assessments was only 36% (direct assessment ICC range for gains: .09- .36). The large differences in ICCs between TS GOLD and the direct assessments show that when children within a classroom are assessed by independent data collectors using the direct assessments they look much less similar to one another than when they are assessed by teachers using TS GOLD.

Many factors may influence the ICCs and certainly families residing in similar neighborhoods are more likely to have similar out-of-care experiences and are likely to cluster within centers. Similar out-of-care experiences (e.g., similar neighborhood experiences, families who have similar cultural, educational, and economic backgrounds,) may mean that children arrive to school with global similarities. These similarities across children might influence teachers' ratings by leading teachers to rate children as more similar to other children within their classroom—resulting in higher ICCs. Whereas, the direct assessments may be less influenced by children's out-of-care experiences and may be better able to capture differences between children across readiness domains—resulting in lower ICCs. However, more rigorous

research is needed that examines the multiple sources of variance at the classroom level that contribute to the ICCs—such as teacher and child characteristics, features of assessments, differences in teacher training and/or classroom composition.

Examining Gains

Our final set of analyses examined the extent to which teacher reports on the TS GOLD are tapping into something meaningful related to children's learning gains as measured through direct assessments. However, again the evidence for discrimination across skills was limited. For example, we found that children's gains on the TS GOLD math domain were significantly predictive of children's skill gains across all directly assessed domains and actually predicted children's gains in directly assessed literacy skills more strongly than children's directly assessed math skills. Examining children's gains in readiness skills from fall to spring extends previous work (Miller-Bains et al., 2017) and the results provide evidence that teachers' assessments of children's readiness skills using TS GOLD are limited in their capacity to measure changes in children's readiness skills in ways that can be differentiated across learning domains.

Implications

Across analyses and time points (i.e., fall, spring, and over the year) we found that teachers' assessments of children's readiness skills using TS GOLD lack precision in discriminating between both children within a classroom and skills within a child. ECE programs have chosen TS GOLD explicitly because it is a comprehensive readiness assessment that can be used longitudinally to provide formative information at the individual child level to help guide instruction within and across learning domains (Lambert et al., 2014). However, TS GOLD may be better suited to providing an overall indication of children's readiness skills and does not accurately measure differences within children across domains or across children within a

classroom. Alternatively, it is possible that the direct assessments create more variability across children within a classroom than actually exists because these assessments are completed out-of-context and do not consider features of the classroom (e.g., engaging learning activities or interactions with peers and adults) that influence how children learn and grow together. In either case, recent recommendations from Sabol and Pianta (2017) suggest that perhaps the prototypic approach to measuring school readiness should be a multi-informant (i.e., observations, teacher-report and direct assessments) assessment because both strengths and areas of needed support often characterize young children's school readiness profiles. However, there is a lack of available literature examining the psychometric properties of school readiness assessments despite recommendations that encourage a "whole child" approach to the assessment of children's readiness skills (Wechsler, Melnick, Maier, & Bishop, 2016). Furthermore, it is critical that school readiness assessments demonstrate sound psychometric properties (reliability and validity) in order to inform appropriate selection of individualized instructional supports. Mandinach (2012) highlights the importance of using the "right" data—meaning data that is aligned to and valid for its intended use as an important factor to help overcome the challenges of data-driven decision making in schools. However, more research is needed before specific assessment recommendations should be made a state-wide level.

Limitations

An important limitation of this study as described above is that the content of the items on TS GOLD are not always perfectly aligned with the items on the direct assessments. The alignment between TS GOLD and the direct assessments was especially weak for the TS GOLD cognitive and social-emotional domains. First, the direct assessments of executive functions measure a narrower skill band (i.e., inhibitory control, working memory and attention) than the TS GOLD domains. Identifying direct assessments that align with the TS GOLD cognitive and

social-emotional domains is challenging given that these domains of TS GOLD capture a range of skills—from managing feelings and making friends to using classification skills and thinking symbolically. Further, because the direct assessments of executive functions in particular measure a narrower skill band, they are somewhat limited in their ability to yield information that is directly applicable to classroom strategies.

Another limitation is that this study was conducted in the first year of teachers' implementation of TS GOLD, and no data were collected on teachers' training on TS GOLD and we were unable to examine Cronbach's alpha for TS GOLD in our sample because we did not have access to the data underlying TS GOLD scores. It is possible that more experienced users of TS GOLD may be better able to differentiate among skills and between students because more exposure to skills across early learning domains may allow them to better observe children and make ratings of skills. However, the present study cannot address this hypothesis. Future work is needed to determine if these results will replicate within a sample of teachers who experience high-quality training and achieve reliability, or who have been using TS GOLD over multiple years. Overall, more research is needed to inform ECE programs using performance-based, observational measures as to how they might improve teacher implementation and validity. However, in this study we did not explicitly examine the factors that may be explaining the lack of evidence of discriminant validity.

There are also some important limitations regarding the sample to be noted. First, as noted above the sampling strategy of selecting one classroom per center, we cannot disentangle if classroom-level effects are due to the individual classroom or the program. Second, the direct assessments were only conducted in English and children had to be proficient enough in their receptive and expressive language skills in English to engage in the direct assessments in order

to be included in the study. The decision to conduct the direct assessments in English only was made due to budget constraints and that less than five percent of the state's pre-K-12 population was classified as English Language Learners. While these decisions limit the generalizability of the results, a strength of this study was that it was conducted within a sample of programs serving a racially diverse population of children coming from economically disadvantaged backgrounds across early childhood education sectors (state-funded preschool, Head Start, private and faith-based centers). Therefore, results of this study provide critical information about a population of young children who are often underrepresented in research.

Conclusions and Future Directions

Understanding young children's foundational school readiness skills is of great interest to teachers, programs, directors, and policy makers looking to improve the quality of ECE experiences in order to support children's development. Currently, there is limited research on the utility of formative assessments in supporting development, yet teachers are being asked to spend a substantial amount of time collecting data. Developmental science does not fully understand the psychometric properties of these assessments, how they are implemented in practice, or how they should be translated into teacher practice and system-wide supports. Understanding this gap is critical given both states' financial investment in assessments and teachers' time investments.

Our study explored the validity of a widely-used, performance-based assessment with a diverse sample of children with the aim of broadening our understanding of these assessments. We found that when compared to direct assessments, TS GOLD showed limited ability to differentiate among children's readiness skills and between children within a classroom both in

the fall, spring and over the course of the year. Importantly, from this study, we cannot speculate on how teachers used or might have used the data for classroom instructional purposes.

These findings raise important questions for the field in terms of research and practice. How much can we confidently say about the utility of performance-based assessments, and of readiness assessment more broadly? ECE programs must clearly delineate their intended purposes for a readiness assessment (i.e., individualization of supports for children, tracking of program level outcomes over time, targeting professional development or curriculum), however once the purposes of the assessments are identified, what available measures are reliable and valid enough for the task at hand? More research is needed in order to understand and refine existing readiness assessments with respect to psychometrics. Only after this is done can research explore how these assessments can and should be used for formative purposes in the classroom and for policy change at the larger systems level.

This paper represents a first step in unpacking underlying efforts to increase formative assessment use in early educational spaces. Until more research is conducted examining the psychometric properties of readiness assessments, the implementation of the assessments, and use of the data, we must be cautious in recommendations that we make to ECE programs, especially when these recommendations are being made at the state or federal level.

References

Ackerman, D.J. (2018). *Real world compromises: Policy and practice impacts of kindergarten entry assessment-related validity and reliability challenges*. Retrieved from: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ets2.12201

Ackerman, D. J., & Coley, R. J. (2012). *State pre-k assessment policies: Issues and status.* Retrieved from https://files.eric.ed.gov/fulltext/ED529449.pdf

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological testing.* Washington, D.C.: American Educational Research Association.

Begeny, J. C., Eckert, T. L., Montarello, S. A., & Storie, M. S. (2008). Teachers' perceptions of students' reading abilities: An examination of the relationship between teachers' judgments and students' performance across a continuum of rating methods. *School Psychology Quarterly*, *23*, 43-55. doi: 10.1037/1045-3830.23.1.43

Bierman, K. L., Domitrovich, C. E., Nix, R. L., Gest, S. D., Welsh, J. A., Greenberg, M. T., ... & Gill, S. (2008). Promoting academic and social‐emotional school readiness: The Head Start REDI program. *Child Development*, *79*, 1802-1817. doi: 10.1111/j.1467-8624.2008.01227.x

Blair, C., & Raver, C. C. (2015). School readiness and self-regulation: A developmental psychobiological approach. *Annual Review of Psychology*, *66*, 711-731. doi: 10.1146/annurev-psych-010814-015221

Boivin, M., & Bierman, K. L. (2014). School readiness: Introduction to a multifaceted and

developmental construct. In M. Boivin & K.L. Bierman (Eds), *Promoting school readiness and early learning: Implications of developmental research for practice* (3-14). New York: Guilford.

Bulotsky-Shearer, R. J., & Fantuzzo, J. W. (2011). Preschool behavior problems in classroom learning situations and literacy outcomes in kindergarten and first grade. *Early Childhood Research Quarterly*, *26*, 61-73. doi: 10.1016/j.ecresq.2010.04.004

Burts, D. C., & Kim, D. H. (2014). The Teaching Strategies GOLD assessment system: Measurement properties and use. *Dialog*, *17*, 122-135. Retrieved from https://journals.uncc.edu/dialog/article/viewFile/170/290

Clements, D. H., Sarama, J., & Germeroth, C. (2016). Learning executive function and early mathematics: Directions of causal relations. *Early Childhood Research Quarterly*, *36*, 79-90. doi: 10.1016/j.ecresq.2015.12.009

Connors-Tadros, L. (2014). *Fast fact: Information and resources on developing state policy on kindergarten entry assessment (KEA)*. Retrieved from http://ceelo.org/wp-content/uploads/2014/02/KEA_Fast_Fact_Feb_11_2014_2.pdf

Daily, S., Burkhauser, M., & Halle, T. (2010). A review of school readiness practices in the states: Early learning guidelines and assessments. Early childhood highlights. *Child Trends*, *1,* 1-12. Retrieved from https://eric.ed.gov/?id=ED510470

Downing S. (2003). Validity: On the meaningful interpretation of assessment data. *Medical Education, 37*, 830–837. doi: 10.1046/j.1365-2923.2003.01594.x

Duncan, G. J., & Magnuson, K. (2011). The nature and impact of early achievement skills,

attention skills, and behavior problems. In G. J. Duncan & R. J. Murnane (Eds.), *Whither opportunity? Rising inequality, schools, and children's life chances* (47–69). New York, NY: Russell Sage Foundation.

Dunn, L. M., & Dunn, D. M. (2007). *Peabody picture vocabulary test, Manual* (4th ed.). Minneapolis, MN: Pearson Assessments.

Early, D. M., Iruka, I. U., Ritchie, S., Barbarin, O. A., Winn, D. M. C., Crawford, G. M., ... & Bryant, D. M. (2010). How do pre-kindergarteners spend their time? Gender, ethnicity, and income as predictors of experiences in pre-kindergarten classrooms. *Early Childhood Research Quarterly*, *25*, 177-193. doi:10.1016/j.ecresq.2009.10.003

Engelhard, G. (1994). Examining rater errors in the assessment of written composition with a many-faceted Rasch model. *Journal of Educational Measurement*, *31*, 93-112. Retrieved from http://www.jstor.org/stable/1435170

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. doi: 10.1016/j.learninstruc.2013.11.003

Friedman-Krauss, A., Barnett, W.S., Weisenfeld, G., Kasmin, R., DiCrecchio, N., & Horowitz, M. (2018). *The state of preschool 2017.* Retrieved from: http://nieer.org/wp-content/uploads/2018/05/State-of-Preschool-2017-Full.5.15.pdf

Garcia, E., & Weiss, E. (2015). Early Education Gaps by Social Class and Race Start US Children Out on Unequal Footing: A Summary of the Major Findings in" Inequalities at the Starting Gate." *Economic Policy Institute*. Retrieved from: https://files.eric.ed.gov/fulltext/ED560364.pdf

Heroman, C., Burts, D. C., Berke, K., & Bickart, T. S. (2010). *Teaching Strategies GOLD® objectives for development & learning: Birth through kindergarten.* Washington, DC: Teaching Strategies.

Houdé, O., Rossi, S., Lubin, A., & Joliot, M. (2010). Mapping numerical processing, reading, and executive functions in the developing brain: An fMRI meta-analysis of 52 studies including 842 children. *Developmental Science*, *13*, 876-885. doi: 10.1111/j.1467-7687.2009.00938.x

Howes, C., Burchinal, M., Pianta, R., Bryant, D., Early, D., Clifford, R., & Barbarin, O. (2008). Ready to learn? Children's pre-academic achievement in pre-kindergarten programs. *Early Childhood Research Quarterly*, *23*(1), 27-50. doi: 10.1016/j.ecresq.2007.05.002

Isaacs, J., Sandstrom, H., Rohacek, M., Lowenstein, C., Healy, O., & Gearing, M. (2015). *How Head Start grantees set and use school readiness goals* (OPRE Report# 2015-12a). Washington, DC: US Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation. Retrieved from: http://www. urban. org/publications/2000087.html.

Janus, M., & Duku, E. (2007). The school entry gap: Socioeconomic, family, and health factors associated with children's school readiness to learn. *Early Education and Development*, *18*(3), 375-403. doi: 10.1080/10409280701610796a

Janus, M., & Offord, D. R. (2007). Development and psychometric properties of the Early Development Instrument (EDI): A measure of children's school readiness. *Canadian Journal of Behavioural Science, 39*(1), 1-22. doi: 10.1037/cjbs2007001

Johnson, D. R., & Young, R. (2011). Toward best practices in analyzing datasets with missing data: Comparisons and recommendations. *Journal of Marriage and Family, 73(5),* 926-945. doi:10.1111/j.1741-3737.2011.00861.x

Kagan, S. L., Moore, E., & Bredekamp, S. (1995). *Reconsidering children's early development and learning: Toward common views and vocabulary.* Darby, PA: DIANE Publishing.

Lambert, R. G., Kim, D. H., & Burts, D. C. (2014). Using teacher ratings to track the growth and development of young children using the Teaching Strategies GOLD® assessment system. *Journal of Psychoeducational Assessment*, *32*, 27-39. doi: 10.1177/0734282913485214

Lambert, R. G., Kim, D. H., & Burts, D. C. (2015). The measurement properties of the Teaching Strategies GOLD® assessment system. *Early Childhood Research Quarterly*, *33*, 49-63. doi: 10.1016/j.ecresq.2015.05.004

Lobman, C., Ryan, S., & McLaughlin, J. (2005). Reconstructing teacher education to prepare qualified preschool teachers: Lessons from New Jersey. *Early Childhood Research & Practice*, *7*, 1-15. Retrieved from https://files.eric.ed.gov/fulltext/EJ1084844.pdf

Lonigan, C. J., Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (2007). *Test of preschool early literacy*. Austin, TX: Pro-Ed.

Mandinach, E.B. (2012). A perfect time for data use: Using data-driven decision making to inform practice. *Educational Psychologist, 47,* 71-85. doi: 10.1080/00461520.2012.667064

McClelland, M. M., Acock, A. C., & Morrison, F. J. (2006). The impact of kindergarten learning-related skills on academic trajectories at the end of elementary school. *Early*

Childhood Research Quarterly, *21*, 471-490. doi: 10.1016/j.ecresq.2006.09.003

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*(4), 947. doi: 10.1037/0012-1649.43.4.947

Meisels, S. J., Bickel, D. D., Nicholson, J., Xue, Y., & Atkins-Burnett, S. (2001). Trusting teachers' judgments: A validity study of a curriculum-embedded performance assessment in kindergarten to grade 3. *American Educational Research Journal*, *38*(1), 73-95. doi: 10.3102/00028312038001073

Miller-Bains, K. L., Russo, J. M., Williford, A. P., DeCoster, J., & Cottone, E. A. (2017). Examining the validity of a multidimensional performance-based assessment at kindergarten entry. *AERA Open*, *3*, 1-16. doi: 10.1177/2332858417706969

National Education Goals Panel (1999). *The national education goals report, building a nation of learners* (Report No. ISBN-0-16-048364-6). Washington, DC: U.S. Government Printing Office. Retrieved from: http://govinfo.library.unt.edu/negp/reports/99rpt.pdf

Ponitz, C. E. C., McClelland, M. M., Jewkes, A. M., Connor, C. M., Farris, C. L., & Morrison, F. J. (2008). Touch your toes! Developing a direct measure of behavioral regulation in early childhood. *Early Childhood Research Quarterly*, *23*, 141-158. doi: 10.1016/j.ecresq.2007.01.004

Ponitz, C. 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- 619. doi: 10.1037/a0015365

Regenstein, E., Connors, M., Romero-Jurado, R., & Weiner, J. (2017). Uses and misuses of

kindergarten readiness assessment results. *The Ounce, 6,* 1-48. Retrieved from http://www.theounce.org/wp-content/uploads/2017/03/PolicyConversationKRA2017.pdf

Regenstein, E., Connors, M. C., Romero-Jurado, R., & Weiner, J. (2018). Effective kindergarten readiness assessments: Influencing policy, informing instruction, and creating joyful classrooms. *YC Young Children*, *73*(1), 36-43. Retrieved from: https://search.proquest.com/openview/b2930cd7bcf672d628307055d04ca521/1?pq-origsite=gscholar&cbl=27755

Riley-Ayers, S. (2014). *Formative assessment: Guidance for early childhood policymakers.* Retrieved from: http://ceelo.org/wp-content/uploads/2014/04/ceelo_policy_report_formative_assessment.pdf

Rimm-Kaufman, S. E., Curby, T. W., Grimm, K. J., Nathanson, L., & Brock, L. L. (2009). The contribution of children's self-regulation and classroom quality to children's adaptive behaviors in the kindergarten classroom. *Developmental Psychology*, *45*, 958-972. doi: 10.1037/a0015861

Sabol, T. J., & Pianta, R. C. (2012). Patterns of school readiness forecast achievement and socioemotional development at the end of elementary school. *Child development*, *83*, 282-299. doi: 10.1111/j.1467-8624. 2011

Sabol, T. J., & Pianta, R. C. (2017). The state of young children in the United States: School readiness. In Votruba,-Drzal, E. & Dearing, E. (Eds.), *Handbook of early childhood development programs, practices, and policies* (3-17). West Sussex, UK: John Wiley & Sons, Inc.

Schilder, D., & Carolan, M. (2014). *State of the states policy snapshot: State early childhood assessment policies.* Retrieved from: http://ceelo.org/wp-

content/uploads/2014/03/CEELO_policy_snapshot_child_assessment_march_2014.pdf

Schmitt, S. A., Pratt, M. E., & McClelland, M. M. (2014). Examining the validity of behavioral self-regulation tools in predicting preschoolers' academic achievement. *Early Education and Development*, *25*, 641-660. doi: 10.1080/10409289.2014.850397

Schrank, F. A., & McGrew, K. S. (2001). Technical abstract (Woodcock-Johnson III assessment service bulletin no. 2). Itasca, IL: Riverside Publishing.

Shields, K. A., Cook, K. D., & Greller, S. (2016). *How kindergarten entry assessments are used in public schools and how they correlate with spring assessments.* Retrieved from https://files.eric.ed.gov/fulltext/ED569203.pdf

Smith-Donald, R., Raver, C. C., Hayes, T., & Richardson, B. (2007). Preliminary construct and concurrent validity of the Preschool Self-regulation Assessment (PSRA) for field-based research. *Early Childhood Research Quarterly*, *22*, 173-187. doi: 10.1016/j.ecresq.2007.01.002

Snow, K. L. (2006). Measuring school readiness: Conceptual and practical considerations. *Early Education and Development*, *17*, 7-41. doi: 10.1207/s15566935eed1701_2

Snow, C., & Van Hemel, S. (2008). *Early childhood assessment: Why, what, and how.* Washington, D.C.: The National Academies Press.

Soderberg J., Stull S., Cummings K., Nolen E., McCutchen D., Joseph G. (2013). *Inter-rater reliability and concurrent validity study of the Washington Kindergarten Inventory of Developing Skills (WaKIDS).* Retrieved from http://depts.washington.edu/cqel/PDFs/CQELReports/WaKIDS_Report072613.pdf

Waterman, C., McDermott, P. A., Fantuzzo, J. W., & Gadsden, V. L. (2012). The matter of assessor variance in early childhood education—Or whose score is it anyway? *Early*

Childhood Research Quarterly, *27*, 46-54. doi: 10.1016/j.ecresq.2011.06.003

Wechsler, M., Melnick, H., Maier, A., & Bishop, J. (2016). *The building blocks of high-quality early childhood education programs.* Palo Alto: Learning Policy Institute.

Weisenfeld, G. (2017). *Assessment tools used in kindergarten entry assessments (KEAS) state scan.* Retrieved from http://ceelo.org/wp-content/uploads/2017/01/ceelo_fast_fact_kea_state_scan_2017_01_for_web.pdf

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*, 226-239. doi:10.1037/a0025361.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson tests of achievement*. Itasca, IL: Riverside Publishing.

Young, R., & Johnson, D. R. (2010, May). Imputing the missing Y's: implications for survey producers and survey users. In Proceedings of the American Association of Public Opinion Research Conference (6242-6248). Retrieved from: https://www.researchgate.net/profile/David_Johnson63/publication/266357732_Imputing_the_Missing_Y%27s_Implications_for_Survey_Producers_and_Survey_Users/links/54c1237f0cf2dd3cb95804bc/Imputing-the-Missing-Ys-Implications-for-Survey-Producers-and-Survey-Users.pdf

Zaninotto, P., & Sacker, A. (2017). Missing data in longitudinal surveys: A comparison of performance of modern techniques. *Journal of Modern Applied Statistical Methods, 16,* 378-402. doi: 10.22237/jmasm/1509495600

Zaslow, M., Tout, K., Halle, T., Whittaker, J. V., & Lavelle, B. (2010). *Toward the Identification of features of effective professional development for early childhood educators.*

Literature review. Retrieved from https://files.eric.ed.gov/fulltext/ED527140.pdf

Table 1
Sample Characteristics

	Frequency (%)	*Mean (SD)*
Male	50.4%	
Age in months (spring)		60.78 (3.70)
Child Ethnicity		
Black/ African-American	71.5%	
White/ Caucasian	20.3%	
Hispanic	2.7%	
Other	6.0%	
Family Income < $15,000	34.3%	
Parent Education		
Less than HS	13.1%	
High School/GED	31.7%	
Some College	32.2%	
Associate's Degree	9.3%	
Bachelor's or More	13.7%	
Other Language Spoken at Home	10.4%	

Note. Data are drawn from a large sample of Louisiana children in 2014-2015, *N* ranges from 816-1109.

Table 2

Descriptives for TS GOLD Domains and Direct Assessments in the Fall and Spring

	Fall			*Spring*		
	N	*Mean (SD)*	*Range Min- Max*	*N*	*Mean (SD)*	*Range Min-Max*
TS GOLD Literacy	834	578.56 (51.59)	296-739	800	694.16 (53.44)	516-800
TS GOLD Language	848	573.43 (64.17)	234-800	816	694.48 (60.13)	442-800
TS GOLD Math	847	577.01 (51.14)	331-764	801	687.59 (53.92)	467-800
TS GOLD Cognitive	847	574.30 (59.36)	259-800	814	705.74 (63.02)	442-800
TS GOLD Social-Emotional	848	582.10 (58.88)	200-800	816	689.80 (53.31)	446-800
TOPEL Print Knowledge	999	96.11 (14.39)	71-135	976	104.08 (13.26)	63-129
TOPEL Phonological Awareness	997	89.13 (14.93)	54-134	971	94.42 (16.86)	54-131
PPVT	1,009	87.70 (16.04)	20-134	988	92.71 (14.58)	24-134
WJ Picture Vocabulary	1,000	99.29 (10.28)	27-137	993	99.68 (9.12)	47-130
WJ Math Reasoning	999	92.41 (13.89)	50-137	991	96.74 (16.20)	39-143
HTKS	1,007	7.67 (12.87)	0-60	987	15.97 (17.19)	0-60
Pencil Tap	1,013	49 (0.34)	0-1	991	71 (0.30)	0-1

Note. Data are drawn from a large sample of Louisiana children in 2014-2015. TS GOLD scores are mean scores ranging from 200-800; each subtest has a slightly different range of scores considered developmentally appropriate for a given age group, for 4-year-olds this range is roughly 600-700 across all domains. Direct assessments are standardized scores with a national mean of 100 and standard deviation of 15. HTKS is a sum score of the total number correct; Pencil Tap is scored as the percent correct. TS GOLD= Teaching Strategies GOLD, TOPEL= Test of Preschool Early Literacy, PPVT= Peabody Picture Vocabulary Test, WJ= Woodcock Johnson, HTKS= Head, Toes, Knees, Shoulders Task.

Table 3
Correlations Between TS GOLD Domains and Direct Assessments: Fall, Spring, and Growth

	DA Language	DA Literacy	DA Math	DA EF
TS GOLD Language				
Fall	**.22*** **	.25***	.26***	.18***
Spring	**.37*** **	.36***	.34***	.30***
Growth	**.12** **	.16***	.07	.12**
TS GOLD Literacy				
Fall	.33***	**.42*** **	.42***	.30***
Spring	.48***	**.53*** **	.49***	.38***
Growth	.18***	**.20*** **	.15***	.12**
TS GOLD Math				
Fall	.27***	.32***	**.38*** **	.19***
Spring	.44***	.44***	**.44*** **	.34***
Growth	.15**	.18***	**.14** **	.13**
TS GOLD Cognitive				
Fall	.20***	.26***	.28***	**.18*** **
Spring	.32***	.30***	.29***	**.23*** **
Growth	.12**	.10*	.06	**.06**
TS GOLD Social-Emotional				
Fall	.07*	.12**	.13***	**.07**
Spring	.26***	.23***	.22***	**.23*** **
Growth	.10*	.05	-.005	**.11** **

Note. Data are drawn from a large sample of Louisiana children in 2014-2015, *N* ranges from 812-1016. DA in column titles stands for "Direct Assessment." *$p < .05$, two-tailed. **$p < .01$, two-tailed. ***$p < .001$, two-tailed. Bolded correlations represent within-domain associations. TS GOLD= Teaching Strategies GOLD. All associations between the TS GOLD and the DAs are reported regarding their respective time point/ gain score, for example, TS GOLD Language Fall is associated with the DA Language in the fall.

Table 4
Proportions of Classroom-Level Variance: Fall, Spring and Growth

	Fall	Spring	Growth
TS GOLD			
Language	.62	.52	.62
Literacy	.50	.48	.63
Math	.51	.51	.66
Cognitive	.57	.63	.71
Social-Emotional	.62	.62	.70
Direct Assessments			
Language	.11	.12	.09
Literacy	.12	.21	.36
Math	.10	.19	.32
Executive Functions	.10	.14	.15

Note. Data are drawn from a large sample of Louisiana children in 2014-2015, *N* ranges from 768-1,016. Variance components were calculated in Stata (Version 14) and range from 0-1. TS GOLD= Teaching Strategies GOLD.

Table 5

Standardized Associations between Each TS GOLD Domain Score and Directly Assessed Growth in Students' Readiness Skills

	Spring Direct Assessment Outcomes			
	Language	Literacy	Math	EF
	Est. (*SE*)	Est. (*SE*)	Est. (*SE*)	Est. (*SE*)
Growth on TS GOLD				
Composite Score	.12** (.07)	.22*** (.05)	.16*** (.04)	.14** (.05)
Language	**.09* (.03)**	.17** (.04)	.11** (.03)	.11** (.04)
Literacy	.13*** (.03)	**.24*** (.04)**	.18*** (.03)	.13** (.05)
Math	.13*** (.03)	.21*** (.04)	**.17*** (.03)**	.13** (.04)
Cognitive	.07* (.03)	.12** (.04)	.10* (.04)	**.06 (.04)**
Social- Emotional	.07* (.03)	.09** (.03)	.06 (.04)	**.11** (.04)**

Note. Data are drawn from a large sample of Louisiana children in 2014-2015, *N*=1109; Data were imputed from all available data using multiply impute in Stata (Version 14). All models accounted for nesting of children within classrooms. All models controlled for children's fall direct assessment scores and children's age, sex, and ethnicity. Growth on each TS GOLD domain was entered into individual models predicting each of the four outcomes; thus, each coefficient and standard error in this table represents results from a separate regression analysis. Bolded scores represent within-domain predictions. Because all variables are standardized, coefficients can be interpreted as effect sizes. TS GOLD= Teaching Strategies GOLD, *SE*= Standard Error.

*$p < .05$, two-tailed. **$p < .01$, two-tailed. ***$p < .001$, two-tailed.

Data Appendix A: Domains and Skills Assessed Across Teaching Strategies (TS) GOLD Domains and Direct Assessments

Data Appendix A: Table 1

Domains and Skills Assessed Across TS GOLD and The Direct Assessments

	Skills Assessed	
	TS GOLD	Direct Assessments
Language	• Listens to and understands increasingly complex language • Uses language to express thoughts and needs • Uses appropriate conversational and other communication skills	• Receptive Language • Expressive Language
Literacy	• Demonstrates phonological awareness • Demonstrates knowledge of the alphabet • Demonstrates knowledge of print and its uses • Comprehends and responds to books and other texts • Demonstrates emergent writing skills	• Phonological Awareness • Print Knowledge
Math	• Uses number concepts and operations • Explores and describes spatial relationships and shapes • Compares and measures • Demonstrates knowledge of patterns	• Ability to analyze and solve math problems • Knowledge of basic math concepts • Number patterns
Social-Emotional	• Regulates own emotions and behaviors • Establishes and sustains positive relationships • Participates cooperatively and constructively in group situations	• Inhibitory Control • Focus and attention • Working memory • Cognitive flexibility
Cognitive	• Demonstrates positive approaches to learning • Remembers and connects experiences • Uses classification skills • Uses symbols and images to represent something not present	

Data Appendix B: Descriptions of the TS GOLD Learning Bands

To complete Teaching Strategies (TS) GOLD, teachers observe children during regular, everyday activities on an on-going basis over the school year and collect samples of student work. An online portfolio is then created for each student. In this portfolio, teachers can save photos, observation notes and/or video clips to assist with their ratings. Teachers then use this information to rate children across 38 objectives and 10 broad areas of development. Many of the objectives include dimensions and examples to help guide teachers' ratings. More specifically, teachers examine individual indicators that each correspond to a designated level/ colored band and make a selection—this process is repeated across the 38 dimensions. For example, an indicator would read, "Manages separations without distress and engages with trusted adults." This indicator corresponds to Objective 2 (Establishes and sustains positive relationships), Level 6 and a yellow and green color band. In Data Appendix A, we also provide the skills assessed in each area of development that was included in this study. As outlined above, these ratings are then translated into colored bands, which correspond to developmental expectations. The colored bands for the areas of development included in this study are provided below.

Data Appendix B: Table 1

TS GOLD Social-Emotional Development and Learning Bands

Age or Class/ Grade	Colored Band	Range of Widely Held Expectations
Birth- 1 year	Red	272-387
1-2 years	Orange	387-494
2-3 years	Yellow	458-546
Preschool 3	Green	516-614
Pre-K 4	Blue	589-690
Kindergarten	Purple	653-795

Data Appendix B: Table 2

TS GOLD Language Development and Learning Bands

Age or Class/ Grade	Colored Band	Range of Widely Held Expectations
Birth- 1 year	Red	312-410
1-2 years	Orange	410-492
2-3 years	Yellow	469-554
Preschool 3	Green	524-639

Pre-K 4	Blue	580-721
Kindergarten	Purple	675-800

Data Appendix B: Table 3

TS GOLD Cognitive Development and Learning Bands

Age or Class/ Grade	Colored Band	Range of Widely Held Expectations
Birth- 1 year	Red	313-377
1-2 years	Orange	377-465
2-3 years	Yellow	453-540
Preschool 3	Green	519-634
Pre-K 4	Blue	591-738
Kindergarten	Purple	679-800

Data Appendix B: Table 4

Literacy Development and Learning Bands

Age or Class/ Grade	Colored Band	Range of Widely Held Expectations
Birth- 1 year	Red	200-375
1-2 years	Orange	408-444
2-3 years	Yellow	470-530
Preschool 3	Green	530-610
Pre-K 4	Blue	572-705
Kindergarten	Purple	644-798

Data Appendix B: Table 5

Mathematics Development and Learning Bands

Age or Class/ Grade	Colored Band	Range of Widely Held Expectations
Birth- 1 year	Red	200-200
1-2 years	Orange	441-491
2-3 years	Yellow	483-557
Preschool 3	Green	545-621
Pre-K 4	Blue	615-712
Kindergarten	Purple	691-800

Data Appendix C: Administration Details for the Direct Assessments

Trained data collectors directly assessed children's language, literacy, mathematics and executive functions. Children were assessed individually during the school day during the fall (September- November) and spring (March-June) in a single session lasting approximately 45 minutes in a quiet location away from classroom distractions. A description of the administration details for each direct assessment is provided in the table below.

Data Appendix C: Table 1
Descriptions of the Administration of the Direct Assessments

School Readiness Domain	Direct Assessment	Description of Administration
Language	• Peabody Picture Vocabulary Test (PPVT)	• Using a flipbook, children were shown four pictures and asked to identify the picture corresponding to a word verbally stated by the examiner.
	• Picture Vocabulary subtest of the Woodcock-Johnson III Tests of Achievement	• Using a flipbook, children were required to verbally name visual pictures presented to them.
Literacy	• Phonological Awareness and Print Knowledge subtests of the Test of Preschool Early Literacy (TOPEL)	• Phonological Awareness: Children are asked to say a word and then say what is left over after dropping specific sounds (elision); and, children are asked to listen to specific sounds and combine them to form a word (blending) • Print Knowledge: Children are asked to point to specific letters, names specific letters, identify letters associated with specific sounds, and say the sounds associated with specific letters

Math	• Woodcock-Johnson III Tests of Achievement	• Applied Problems: Children are read story problems and they can follow along in the test booklet. There is no multiple-choice response set. • Quantitative Concepts: Children are asked a variety of questions to assess their knowledge of math concepts (e.g., numbers, shapes, symbols). Some questions provide response options and others are open-ended.
Executive Functions	• Head, Toes, Knees, Shoulders (HTKS) Task	• Children are asked to play a game in which they must do the opposite of what the examiner says. For example, if the examiner says touch your head, the child must touch their toes.
	• Pencil Tap	• Children are required to hold two rules in mind. They must tap their pencil once when the examiner taps theirs twice and twice when the examiner taps theirs once.

Advancements in Understanding How Children's Experiences in Neighborhoods and Classrooms Contribute to Self-Regulation

Jaclyn M. Russo [1]
Amanda P. Williford [1]
Jason T. Downer[1]

[1]University of Virginia, Center for Advanced Study of Teaching and Learning (CASTL)

Abstract

The present study aimed to better understand how children's experiences across contexts (school and neighborhoods) combine to influence the development of self-regulation skills. Specifically, we explored the implementation of a novel, virtual neighborhood coding scheme in order to better capture the physical features and resources that are available in young children's proximal neighborhoods. And, how these proximal neighborhood features combine with children's individual experiences with teachers in the classroom to impact self-regulation development. In a sample of 380 preschool children (mean age= 52.51 months, SD= 3.72 months) in 51 classrooms, we compared features of neighborhoods as assessed by a virtual neighborhood coding scheme to census tract neighborhood data. We then examined the association between these different assessments of neighborhood features and children's growth in self-regulation skills during preschool. Findings indicated some initial evidence for the reliability and validity of the neighborhood coding scheme in the present sample. We also found that for children who resided in neighborhoods with few resources to meet basic and daily needs (i.e., grocery store, doctor's office, pharmacy) that the experience of high-quality, warm and supportive individual interactions with teachers was protective and promoted the growth of self-regulation skills. This study advances the current literature through the use of a novel, virtual tool to better understand children's every day lived experiences in their neighborhoods. And, offers promise for future research that is aimed at understanding strengths and risk factors in communities to inform policy change and neighborhood development that best supports young children's development.

Keywords: *neighborhoods, self-regulation, teacher-child interactions, cumulative stress*

Advancements in Understanding How Children's Experiences in Neighborhoods and Classrooms Contribute to Self-Regulation

Young children's self-regulation skills are foundational not only for school readiness but for later success in school and life (Blair et al., 2011; Blair & Raver, 2015; Duncan, Schmitt, Burke, & McClelland, 2018). Children's experiences across contexts (home, school, community) constrain or support their self-regulation development well before they arrive in kindergarten (e.g., Brown et al., 2013). Given the burgeoning research on the importance of children's early experiences for later development and learning, there has been a recent shift in early childhood education policy to broaden the understanding of young children's readiness for school to include aspects of families, schools, and neighborhoods that can hinder or bolster children's early development (e.g., Coulton, Richter, Kim, Fischer, & Cho, 2016; U.S. Department of Health and Human Services, 2016).

However, some current research misses important variation in children's individual experiences either by focusing on a single context (i.e., school or home) or by using global measures – like census tract data or classroom-level teacher-child interactions – to capture complex phenomena, such as social-emotional development or school readiness (e.g., McCoy et al., 2015; Minh, Muhajarine, Janus, Brownell, & Guhn, 2017). For example, a more nuanced understanding of the resources and risk factors present in children's proximal neighborhoods is paramount, because these features are critical for self-regulation development. Access to green space or a park provides young children with a space to play creatively and actively with their peers, which can reduce stress and its detrimental effects on cognitive systems responsible for the development of self-regulation (Blair & Raver, 2012; Flouri et al., 2014; Taylor, Kuo, & Sullivan, 2002). Simlarly, having easily accessible resources, such as grocery stores or doctors' offices, can help to reduce stress that caregivers feel, which in turn may help caregivers to be

more emotionally available, which then positively impacts parenting behaviors and children's self-regulation development (Davis, Bilms, & Suveg, 2017). Additionally, examination of the quality of young children's individual interactions with their teacher compared to more commonly used global indicators of teacher-child interactions at the classroom-level provides a better understanding of individual children's experiences in the classroom. This is important because children's experiences in the classroom with a consistent, reliable, and emotionally attuned adult can influence growth in self-regulation by providing a secure base for children to access when they are upset or frustrated (Williford et al., 2013). These interactions provide young children with feedback and support that encourages the development of their self-regulation skills. Moreover, these individual interactions may be the most important for children who reside in neighborhoods with many risk factors, because the experience of warm, emotionally attuned interactions with a teacher can reduce feelings of stress experienced outside of the classroom, and support self-regulation development (McCoy et al., 2015). In other words, high- quality individual teacher-child interactions may be capable of providing a protective, moderating effect in regards to self-regulation development for children most at risk. Taken together, the current study leverages the use of novel assessments of children's proximal neighborhoods and teacher-child interactions in order to better understand children's individual experiences that influence their self-regulation development.

Children's Early Self-Regulation Development

Definitions of self-regulation vary but most agree that it includes young children's ability to focus their attention, temper strong emotions and control their behaviors (Campbell et al., 2016; Daily, Burkhauser, & Halle, 2010; McClelland & Cameron, 2012; National Education Goals Panel, 1995; U.S. DHHS, 2015; Williford, Whittaker, Vitiello & Downer, 2013). The

preschool period is a time characterized by the rapid development of self-regulatory skills (Bronson, 2000; McClelland & Cameron, 2012). Children's self-regulation skills develop within context—in school for example, children need to be able to be able to inhibit impulses, sit quietly and filter out distractions during circle time in order to successfully participate in the learning activity. And, aspects of children's early environments can either hinder or support this development (e.g., Campbell et al., 2016).

Young children's experiences across environments and self-regulation development

Young children who experience multiple stressors across environments are more likely to lag behind their peers in critical early developmental areas such as social and emotional and academic domains. This has given rise to the conceptualization of risk factors as cumulative or additive, and this is especially important given that children and families with lower socioeconomic status are more likely to be embedded within environments with multiple stressors (e.g., Blair et al., 2011; Evans, 2003; Evans & Kim, 2013; Vernon-Feagans & Cox, 2013; Garmezy & Rutter, 1983). Further, the experience of multiple stressors across environments is particularly damaging to the cognitive systems responsible for the development and implementation of self-regulation skills (McEwen, 1998). Examination of experiences across children's early environments offers insight into the intra-individual differences of children and families—for example, children may reside in neighborhoods with many risk factors, but may experience warm, supportive interactions with their teacher in the classroom (Ackerman, Izard, Schoff, Youngstrom, & Kogos, 1999). To best capture the nuances that are inherent between children's experiences across environments, Evans and colleagues (2013) asserted the need for the field to examine children's experiences, including supports and resources across multiple environments (i.e., neighborhood and classroom). The presence of social and physical stressors

across contexts leads to chronic physiological stress (e.g., elevated blood pressure and blood cortisol levels) which predicts morbidity into adulthood despite attaining a higher social class later in life (Evans & Kim, 2013). However, identification of strengths and resources as well as risk factors in children's early environments offers a conduit through which early and effective supports might be most impactful for children's development.

Resources and risk factors in children's neighborhoods. Current literature examining the strength of the relationship between features of children's neighborhoods and their self-regulation development is mixed. Some findings indicate that risk factors present in neighborhoods are significantly and negatively associated with child outcomes over and above family factors, while others indicate that the effect is less strong once family factors are taken into account (Heberle, Thomas, Wagmiller, Briggs-Gowan, & Carter, 2014; Vaden-Kiernan et al., 2010). In addition, specific mediation and moderation pathways between risk factors and resources in children's neighborhoods and their self-regulation development have not been clearly established (Minh et al., 2017). Without a more specific understanding of the mechanisms behind these associations it is not possible to provide families, schools, communities or policymakers with meaningful guidance to help improve areas of challenge in neighborhoods. One possible explanation for the mixed findings is that current research relies primarily on global—most often census tract— data to capture children's individual neighborhood experiences (Minh et al., 2017). The problem with this method is that the geographic sizes of census tracts vary widely (U.S. Census Bureau, 2015) and the experiences of children who live within the same census tract can vary significantly. So, it is perhaps not surprising that research using census tract data have yielded inconsistent findings with regards to child outcomes. However, burgeoning advancements in technology (i.e., use of Google Earth) provide both cost

effective and more precise measurement of the features of children's proximal neighborhoods. This will allow for a more robust understanding of how resources and risk factors in neighborhoods contribute to self-regulation development and how to best mobilize supports for children and families.

Additionally, prior research has often relied on aggregate indicators of income or percentage of single- parent households at census tract levels to quantify disadvantaged versus affluent neighborhoods (Caughy et al., 2013; Minh et al., 2017). This again masks important variation within census tracts— and assumes that children's experiences within a census tract are the same. While these studies often find a significant association between concentrated disadvantage and children's social and emotional functioning, the explanatory pathways are unclear and therefore implications for community intervention planning are limited. Given that the majority of children's activities take place within 10 minutes or .5 miles of their homes, census tract data are an imprecise representation of individual children's experiences (Jones et al., 2009). In other words, children and families are more likely to access resources such as parks or libraries when they are in close proximity to their home.

Access to green space, such as parks or playgrounds, predicts children's level of physical activeness, and research suggests that children are most active close to their homes (Jones et al., 2009). Physical activity reduces cortisol levels and feelings of stress—two critical factors that impact the cognitive systems responsible for self- regulation (Blair & Raver, 2012). Some children within a census tract may live in closer proximity to a park, allowing them to more frequently access this space—variation that can only be captured through examination of the space immediately surrounding a child's home.

Similarly, risk factors need to be examined not at a global level but within children's

proximal neighborhoods. A child's everyday individual experiences matter most for their self-regulation development. For example, is crossing the street stressful because there are not adequate safety features? Can a child's caregiver easily obtain fresh food and meet basic needs or is it stressful to find a grocery store or pharmacy? Is there a library or community space nearby for peer interactions that lead to a sense of community cohesion? (Cutrona, Wallance, & Wessner, 2006; Franco, Pottick, & Huang, 2010; Scott, 2011).

The resources (time, transportation) needed for a caregiver to travel several miles to access these opportunities are a barrier that is not captured by census tract level neighborhood data. A thorough literature review revealed one study that examined the physical features (e.g., housing conditions, peeling paint, boarded windows, unkempt lawns) of children's proximal neighborhoods and found that these features were significantly associated with children's behavior problems; importantly, the relevant geographic area was a radius of 400 to 800 meters surrounding a child's home (Caughy et al., 2013). This finding suggests the importance of capturing the features of children's proximal neighborhoods. Furthermore, it is imperative to capture both resources (parks, libraries, fresh food stores, access to public transportation) and risk factors (boarded up homes, trash, lack of street safety) present in children's proximal neighborhoods to accurately and comprehensively capture their experiences—the current study leverages a new measurement tool, using Google Earth, that will allow us to fill this gap. With this level of individualized measurement of the features of children's neighborhoods, we will be better equipped to understand how these neighborhood features interact with children's individual experiences in another important context—the classroom, with their teachers—and to provide effective and individualized supports.

Teacher-Child Interactions in the Preschool Classroom. Children's interactions

with their teachers play a formative role in their social-emotional development (e.g., Eisenberg, Valiente, & Eggum, 2010; Hamre & Pianta, 2005; O'Connor & McCartney, 2007; Pianta & Hamre, 2009; Sabol & Pianta, 2012). This is especially true of early foundational skills, such as self-regulation, as young children are now in an environment where they are required to navigate interactions with adults and their peers (Rimm-Kaufman et al., 2009). Therefore, the experience of high quality, sensitive interactions with teachers in the classroom is well positioned to support the development of self-regulation skills. For example, the mission of Head Start is to provide young children with the experience of high-quality interactions with a stable adult that they can rely on in the classroom—their teacher(s). These experiences in the classroom are especially important for children who reside in neighborhoods with many risk factors, because their early self-regulation skills are likely to be underdeveloped when they arrive to school. However, most current research has relied on global, classroom-level indicators of teacher-child interaction quality (e.g., Cadima, Verschueren, Leal, & Guedes, 2016; Hamre, Hatfield, Pianta, & Jamil, 2014; Hamre & Pianta, 2005; Rimm-Kaufman, LaParo, Downer, & Pianta, 2005). Similar to census tract estimates of neighborhood quality, classroom-level teacher- child interaction quality measures mask important variation in children's individual experiences within a classroom.

Teachers who are more attuned to or sensitive to *individual* children's emotional needs are better able to provide them with *individualized* effective supports (e.g., a quiet space to calm down in the classroom when they are becoming upset) to encourage their developing self- regulation skills (Hamre & Pianta, 2001, 2005; Rimm-Kaufman et al., 2009). However, not all children may have the same experiences with their teacher in the classroom—variation that is only captured through examination of the quality of *individual* teacher-child

interactions. In addition to teacher sensitivity, individual teacher-child interactions that are warm and characterized by open communication allow children to feel comfortable approaching their teacher and talking about their feelings, using the teacher as a source of support when upset (Buyse, Verschueren, & Doumen, 2011; Pianta & Stuhlman, 2004; Sabol & Pianta, 2012). These features of early individual teacher-child interactions are particularly influential in supporting children's self-regulation development, as teachers can serve a regulatory role, providing a secure emotional base for children in the classroom (Buyse et al., 2011, Williford et al., 2013). When children experience warmth, trust, and low levels of conflict in their interactions with their teacher, they show better adjustment to school, more positive school outcomes (both behavioral and academic), and fewer externalizing problems (Baker, Grant, & Morlock, 2008; Pianta & Stuhlman, 2004). In a sample of Head Start children, the quality of teacher-child interactions mediated the relationship between neighborhood economic disadvantage and behavior problems—such that higher levels of negative teacher-child interactions predicted increases in behavior problems (McCoy et al., 2015).

High-quality teacher-child interactions directly influence children's developing self-regulation skills (e.g., Hamre & Pianta, 2001; Williford et al., 2013). However, these interactions are particularly important for children who experience stressors outside of the classroom, because they are more likely to enter preschool and kindergarten with underdeveloped self-regulation skills compared to their peers who do not experience many stressors outside of the classroom—and the experience of a supportive and warm relationship with a teacher can reduce stress and support self-regulatory abilities. In other words, high-quality teacher-child interactions can *moderate* the relationship between neighborhood

resources and risk factors and self-regulation development. However, a major gap in the current literature is that most research to date has examined this relationship based upon classroom-level averages of teacher-child interactions. Williford et al. (2013) found that the quality of individual teacher- child interactions predicted growth in children's self-regulation skills, offering promising evidence for the importance of examining individual teacher-child interactions in relation to child outcomes over and above classroom-level indicators of teacher-child interaction quality. More research is needed to better understand the links between *individual* teacher-child interaction quality and how these interactions might moderate the relationship between resources and risk factors in *proximal* neighborhoods and self-regulation development for children who are most at risk. The present study aims to fill this gap.

Present Study

The purpose of the present study was to better understand how the features (availability of resources, safety features, physical order) of young children's neighborhoods combines with the experiences that they have with teachers in the classroom to influence growth in their self-regulation skills. The novel measurement of children's experiences in neighborhoods (virtual neighborhood observation coding) and in the classroom (quality of individual teacher-child interactions) provides increased specificity with regards to individual children's experiences. We examined the direct links of the virtual neighborhood observation coding to growth in children's self-regulation skills during preschool, beyond more commonly used global neighborhood indicators. In addition, we explored how children's individual experiences in their *proximal* neighborhoods interacts with their *individual* experiences with their teacher in the classroom to influence growth in their self-regulation skills during

preschool. Given the novelty of the virtual coding scheme, we also explored and described the descriptive properties of the neighborhood coding composites in addition to the primary research questions. Specific research questions and hypotheses are described below.

Research Question 1

Can the virtual neighborhood coding scheme be applied to the present sample of children's neighborhoods with sound reliability and validity? And, what are the descriptive features of children's neighborhoods in the present sample? We explored whether the virtual neighborhood coding scheme could be reliably and validly applied to a sample of children from suburban, rural and urban clusters (outside of an urbanized area with at least 2,500 inhabitants; Ratcliffe et al., 2016) in one southeastern U.S. state compared to the validation sample (McCoy et al., 2019) of children who were from suburban, urban clusters, and large urbanized areas across the US. Given that children's communities in the present sample were less densely populated compared to the children's communities in the validation sample, we expected that there would be fewer resources present per proximal neighborhood. In addition, certain features of the coding scheme (i.e., bars on windows, graffiti) are typically more common in urbanized areas and we therefore also expected limited variability in some of the codes that assessed the physical signs of order in the neighborhood.

Research Question 2

Does assessing children's proximal neighborhoods provide additional information over and above assessment of neighborhoods using census tract level data? Given that the virtual neighborhood observation coding scheme captures information about children's proximal neighborhoods (within a .5 mile radius), we expected that the coding of the resources and risk factors present in children's proximal neighborhoods would provide information about aspects

of children's neighborhoods that is more nuanced and descriptive (i.e., a community center down the block) than what is available at the census tract level, which has broader boundaries—thereby capturing more variability between features of children's proximal neighborhoods. We expected that most of the variance in children's environments would be accounted for not by their census tract membership but by the boundaries immediately surrounding their homes.

Research Question 3

After accounting for features of children's homes and families, does the virtual neighborhood observation coding predict growth in children's self-regulation skills during preschool, over and above census tract level neighborhood indicators? Consistent with prior research, we expected that adverse neighborhood experiences would be associated with less growth in self-regulation skills over the preschool year compared to children who reside in neighborhoods with fewer risk factors (Blair et al., 2011; Evans, 2003). The virtual neighborhood observation coding scheme advances the current literature by providing a more precise measurement of children's everyday experiences in their neighborhood. This enhanced specificity is critical to understanding how self-regulation develops in early childhood. Census tract level data does not offer this level of specificity. Therefore, we hypothesized that after accounting for children's experiences at home, the virtual neighborhood observation coding scheme would predict growth in children's self-regulation skills over and above census tract level data.

Research Question 4

Does the relationship between resources and risk factors in children's neighborhoods [as measured by the virtual neighborhood observation coding scheme] and children's growth

in their self-regulation skills differ depending upon the quality of individual teacher-child interactions experienced in a preschool classroom? We expected the quality of children's individual interactions with their teachers in the classroom to moderate the relationship between resources and risk factors present in their neighborhoods as measured through the virtual neighborhood observation coding and their growth in self-regulation skills. We hypothesized the positive association between children's neighborhood resources and the development of their self-regulation skills to be enhanced as the quality of their individual interactions with their teacher increased. We also expected children's individual high-quality teacher-child interactions to serve as a protective factor, such that the negative association between neighborhood risks and children's gains in self- regulation would be weakened as the quality of individual teacher-child interactions was greater.

Method

Participants

Data for this study is leveraged from the P2K (Understanding the Power of Preschool for Kindergarten Success) project, Preschool Year, Cohort 1. Data were collected on 380 children and 53 teachers from 51 preschool classrooms that included both Head Start and state-funded preschool programs across two school divisions in Virginia. Children in this study are ethnically diverse (59.1% Black, 32.1% White, 8.8% Hispanic), and 34.2% of the families earn less than $20,000 annually. The mean age of the children at the start of the school year was 52.51 months (*SD*= 3.72) and 49.9% were male. The majority of teachers in this study were female (98%) and white (69.8%) with a mean age of 44.59 years (SD= 10.43 years) and over half (54.9%) possessed a master's degree. See Table 1 for participant characteristics.

Data Collection Procedures

Recruitment. All study procedures were approved by the university IRB. State-funded preschool programs serving primarily low-income families were eligible and classrooms within programs were deemed eligible if they served primarily 4-year-olds. Specifically, two districts within one southern state agreed to participate and classrooms were recruited from within those two divisions. Teacher consent packets were distributed to all teachers in eligible classrooms. A total of 53 teachers participated in the study—this number is slightly higher than the total number of classrooms and reflects changes that occurred over the year (e.g., one teacher was on maternity leave and an assistant participated in her place until her return, another teacher was on medical leave in the spring and again her assistant agreed to participate).

Children were eligible to participate in the study if they were 4-years-old on or before September 30, 2016. Children with active IEPs and those who spoke a language other than English were eligible to participate as long as they were able to complete the assessments in English. Once a teacher agreed to participate in the study, the study team distributed consent packets for all eligible children in their class and these packets were distributed during the first week of school. Once consent packets were returned, 4 boys and 4 girls were randomly selected within each classroom. Random selection occurred first within the classrooms with the highest number of returned consents and last within the classrooms with the fewest number of returned consents to ensure that as many classrooms as possible had at least eight consented children. There was a range in the number of children whose parents consented to participation within each classroom (range: 4-16 children per classroom), and on average 9 children (SD= 2.64) were consented for participation per classroom. If a child who was selected to participate in the fall was unable to participate in the spring, a replacement child

from the same classroom was selected. In total, 380 children participated in the fall and 354 in the spring (9% attrition). Reasons for child attrition include, that the child moved out of area, could not be assessed due to a disability or other specified reason.

Data collection. Parents and teachers completed surveys and data collectors conducted classroom observations, and direct assessments of children's skills. Local data collectors were hired to conduct observations and direct assessments and to facilitate collection of parent and teacher surveys. Data collectors serving primarily as direct assessors completed an orientation and 3-day training, which totaled 24 hours of training. They practiced each direct assessment during the training and were encouraged to contact study staff should any questions arise during data collection. Data collectors conducting observational data completed a 5-day comprehensive training (40 hours total). Observational data collectors achieved reliability on the observational measures—becoming certified observers. In addition, observation coding calibration meetings occurred five times over the school year to ensure continued reliability.

Parent surveys were distributed in conjunction with the child consent and parents had the option of completing a paper copy or entering their survey responses online. All parents received a $10 gift card for completion of the consent and survey and one family received an additional $100 gift card following a random drawing. Direct assessments were administered to children by trained data collectors during the school day in the fall (September 26- November 10, 2016), winter (January 9- February 17, 2017) and the spring (April 17- May 26, 2017). Data collectors obtained assent from the children prior to testing and were trained to monitor for fatigue and offer breaks as needed. Classroom observations were completed in teams of two (approximately 20% of observations were double-coded) over three days in the fall, winter and spring—during the same windows as the direct assessment data collection. Observations ranged across

classroom activities, excluding naptime. The neighborhood coding data was collected by a team of undergraduate and graduate coders under the guidance of the first author. Details regarding the data collection procedures for this tool are described in more detail in the measures section below.

Measures

Family and Home Covariates

Characteristics of children's homes and families that are associated with the development of children's early self-regulation skills were captured through primary caregiver report. The three measures are described below.

Household chaos. Household chaos was measured through six items from the *Confusion, Hubbub, and Order Scale* (CHAOS; Matheny, Wachs, Ludwig, & Phillips, 1995). Items were developed to depict household characteristics that reflect chaotic environments (e.g., "It's a real zoo in our home" or "You can't hear yourself think in our home"). Items are scored on a five-point scale (1= Definitely untrue; 5= Definitely true). Items 1, 4 and 6 are reflective of routines and order (e.g., "The children have a regular bedtime routine") and were reversed scored to be on the same scale as the other items, reflecting higher levels of chaos and disorder. The final score is the raw sum of total possible points (6 items each with a five-point scale = 30 total possible points). The CHAOS shows evidence of internal consistency reliability (Cronbach's alpha = 0.79) and construct validity in previous samples. Specifically, the CHAOS was significantly associated with live observations of the conditions of families' homes (Matheny et al., 1995). In the present sample, the CHAOS demonstrated moderate internal consistency (Cronbach's alpha= .60).

Parental Distress. Parental distress was measured through the Kessler Psychological

Distress Scale—10-question or K10 version (Kessler et al., 2002). The K10 is a short dimensional measure of non-specific psychological distress (anxiety and depressive symptoms). Items are scored on a five-point scale (1= None of the time; 5= All of the time). The final score is the raw sum of total possible points (10 items each with a 5-point scale = 50 total possible points). In a large, representative sample, item response theory (IRT) analysis was used to determine precision of the scale in discriminating between clinical and non-clinical samples – results showed good precision in the 90-99th percentile range (Kessler et al., 2002). The K10 also demonstrates strong internal consistency in previous literature (Cronbach's alpha= .93; Fassaert et al., 2009) and in the current sample (Cronbach's alpha=.86).

Classroom-Level Teacher-Child Interaction Quality Covariate

At the classroom-level, we controlled for classroom level quality of teacher-child interactions using the *Classroom Assessment Scoring System* (CLASS; Pianta, La Paro, & Hamre, 2008). The CLASS captures aspects of the quality of teacher interactions (across three domains: Emotional Support, Classroom Organization, Instructional Support) experienced by most children within a classroom (Pianta et al., 2008). For the present study, teachers' scores on the CLASS Emotional Support domain were aggregated across the three time points over the year (fall, winter, and spring) to provide an average across the year of the quality of emotional support at the classroom level. Over the course of the school year, 19.9% of all cycles were double-coded and the Emotional Support domain demonstrated strong inter-rater agreement (ICC= .90; Cronbach's alpha= .87).

Demographic Covariates

The following parent and child covariates were included to increase the precision of our

estimation. For parents, the following covariates were included in all analyses: income to needs ratio (coded dichotomously; 1= below the federal poverty line and is based on income and number of persons in the household), and parents' educational attainment (coded dichotomously; 1= at least an associates degree, which is the indicator of higher education in the Commonwealth of Virginia). For children, the following covariates were included in all analyses: sex (1= boy) and race/ethnicity (African American, Caucasian, Hispanic; categories are not mutually exclusive).

Neighborhood Characteristics

Resources and risk factors present in children's neighborhoods were captured in two ways—first, a virtual neighborhood observation coding scheme captured several features of children's proximal neighborhoods and second, census tract level information about crime levels and housing quality was captured through a state-wide index of youth well-being, The Youth Well-Being Index. The two measures are described below.

Virtual neighborhood observation coding. The virtual neighborhood observation coding scheme used in this study was an adapted version of a recently developed and validated tool, the Virtual Systematic Social Observation- Tallying observations in urban regions (SSO i-Tour; Odgers, Caspi, Bates, Sampson, & Moffitt, 2012). The original tool was developed to capture positive and negative neighborhood features by leveraging advancements in technology (i.e., Google Street View and Google Earth) to adapt in-person SSO measures for a virtual context (Odgers et al., 2012). The original SSO i-Tour contains 18 total items and has been applied with moderate to strong inter-rater reliability (ICC coefficients ranging from .48 to .91; Odgers et al., 2012). Additionally, convergent and discriminant validity were established through the examination of associations between the virtual SSO i-Tour and census-based neighborhood

information. For example, census-classified neighborhood deprivation was associated with more negative neighborhood features as assessed by the virtual SSO i-Tour (*r*'s range from .30 to .47). Further, negative features of neighborhoods as captured by the virtual SSO i-Tour were significantly predictive of children's antisocial behavior but not children's pro-social behavior (Odgers et al., 2012).

In collaboration with the authors of the SSO i-Tour, McCoy and colleagues (2019) adapted the original protocol to increase the breadth of neighborhood features captured and align more closely with characteristics present in U.S. neighborhoods (e.g., street signage, modes of public transportation, types of food stores). The adapted observation coding protocol was designed to code the proximal neighborhood surrounding children's schools—the current study used this more comprehensive protocol to code the neighborhood surrounding children's homes. The adapted virtual neighborhood observation protocol contains 50 total items that capture the physical signs of order (e.g., quality of houses and buildings, presence of trash), safety (e.g., presence of sidewalks and crosswalks) and resources (e.g., parks, libraries, food stores, doctors' offices) present in neighborhoods. Following the recommendations of Odgers and colleagues (2012), Google Street View was used to locate children's home addresses and create boundaries using a .5-mile radius to define children's proximal neighborhoods. Then, using tools available through Google Earth (a free downloadable program), boundaries were created for each child's proximal neighborhood thereby creating a separate Google map for each child.

After all maps were created, individual children's neighborhood maps were randomly assigned to two coders. All maps were double-coded. All identifying data was removed from the maps including children's home addresses to ensure confidentiality. Coders completed the

coding on their own personal machines. Coders used the Google Street View option within Google Earth to virtually "walk" children's neighborhoods while simultaneously entering codes into a Qualtrics survey. It took on average between 20-30 minutes to code a single neighborhood. Weekly calibration meetings were held with the entire coding team and the lead author to ensure reliability. All final codes were on 0-1 scale, with higher scores reflective of more positive neighborhood features (e.g., higher quality houses or the presence of a resource).

YWBI (Youth Well-Being Index). Two variables from the Virginia Youth Well-Being Index (YWBI) were used—a neighborhood crime indicator and a neighborhood housing indicator. The YWBI is linked at the census tract level—census tracts are geographic areas identified statistically by the U.S. Census Bureau that are meant to reflect children's neighborhoods. We used children's home addresses to link them to a census tract, which was then linked to the corresponding YWBI indicators for that census tract. The crime indicator assesses crime levels using data from the Federal Bureau of Investigations Uniform Crime Report. The housing indicator assesses housing quality, including the proportion of housing units built before 1951 and without plumbing facilities, along with the proportion of overcrowded households defined as more than two occupants per room. Across both school divisions in this study, approximately 28% of the population under age 18 are living in census tracts that are classified as "Very Low Opportunity" (Virginia Department of Health, 2017). Both variables are standardized, and range from 0-1. Higher scores reflect less disadvantage.

Quality of Individual Teacher-Child Interactions

The quality of an individual child's interactions with their teacher was measured through the *Individualized Classroom Assessment Scoring System* (inCLASS; Downer, Booren, Lima, Luckner, & Pianta, 2010) at three points over the preschool year (fall, winter, and spring). The

inCLASS measures a child's interactions with their teachers, their peers, and tasks. For this study, only the Teacher Interactions domain was used—this domain of the inCLASS is comprised of two dimensions: Positive Engagement with the Teacher and Teacher Communication. Each dimension is scored on a 7-point scale based on the observation of specific behavioral markers—with higher scores indicating higher quality teacher-child interactions. Each inCLASS observation cycle lasts 15 minutes (10 minutes of observation and 5 minutes of scoring). During each of the three time points throughout the preschool year, observers completed approximately six inCLASS cycles for each child sequentially (approximately 1.5 hours per child) and spent at least three days in the classroom to complete all observations for all consented children. Our final score is an aggregate of all cycles across the three time points. The Teacher Interactions domain of inCLASS demonstrates strong inter-rater reliability in past research (ICC = .83; Downer et al., 2010). The Teacher Interactions domain is also positively associated with teacher ratings of closeness with a child and teacher reports of assertiveness—providing evidence of criterion-related validity (Downer et al., 2010). In the present sample, over the course of the school year, 19.9% of all cycles were double-coded and the Teacher Interactions domain demonstrated strong inter-rater agreement (ICC= .87; Cronbach's alpha= .74).

Direct Assessment of Children's Self-Regulation

Children's behavioral self-regulation was directly assessed using the Head-Toes-Knees-Shoulders (HTKS) task (Ponitz et al., 2008). HTKS was administered in the fall, winter, and spring of the preschool year. HTKS measures inhibitory control, working memory, and attention by asking the child to do the opposite of what the examiner says (e.g., touch your head when I say "Touch your toes"; Ponitz et al., 2008). The total score is the sum of correct items (range: 0-

60). The HTKS shows evidence of strong concurrent and construct validity—significant associations with teacher-rated self-regulation (*r*'s range from 0.15- 0.20, all *p*'s < .05)—and inter-rater reliability (alpha= 0.95; Ponitz et al., 2008). Within the present sample, there was moderate association (*r*'s ranged from .53-.63) across the three timepoints.

Analytic Strategy

Missing Data

Of the full sample of 380 children, not all had full and complete data across all variables. With regards to the neighborhood predictors, 84% of children had complete data on the virtual neighborhood coding composites and 82% had complete data on the YWBI. For the classroom-level observational data, 99% of children had their classroom observed in the fall and winter and 94% had their classroom observed in the spring. For the individualized observation of teacher-child interaction quality, 99% of children were observed in the fall, 98% in the winter and 93% in the spring. On the direct assessment of self-regulation (i.e., HTKS), 99% of children were assessed in the fall, 96% in the winter, and 93% in the spring. With regards to demographic characteristics, 11% of children were missing information about their identified racial group, and all children had complete data on their sex identification and age.

To address missingness we first compared the covariate information between those children who were missing data and those children who had complete data. T-tests across all child- and family-level demographic characteristics, including age, race, sex, family income, parent education, household chaos and parental distress, indicated no significant differences (all t-test *p* values >.05) between children missing data and those without missingness.

To maximize all available data from the sample, all descriptive analyses (including means, standard deviations, correlations and intraclass correlations) were calculated using the

largest possible sample. The largest possible sample was obtained through listwise deletion. Missing data in regression models was accounted for using full information maximum likelihood estimation with robust standard errors to make use of all available data for predictor variables.

Data Analyses

Descriptive information and correlations were computed in SPSS Version 25. Regression models examining the moderation of children's experiences with their teacher in the classroom on the relation between neighborhood characteristics and growth in self-regulation were completed using MPlus version 8.1 (Muthen & Muthen, 2018). The Type=Complex function was used to adjust standard errors to take into account that children were clustered within classrooms.

Research question 1. For this first research question, we were interested in whether or not the virtual neighborhood coding scheme could be applied to the present sample of children's neighborhoods with sound reliability and validity. We were also interested in better understanding the descriptive features of the present sample of children's neighborhoods.

All neighborhoods were coded independently by two coders and inter-rater reliability (IRR) was calculated for individual codes using Kappa and intraclass correlation coefficient (ICC) statistics (Hallgren, 2012). More specifically, following the recommendations of Hallgren (2012), the Kappa statistic was used to examine IRR for all categorical codes and the ICC was used to examine IRR for all ordinal and continuous codes. Double-codes were then averaged to create a single code for every individual item within the coding scheme. For example, if Coder 1 coded grocery store as 1/Present and Coder 2, for the same neighborhood, coded grocery store as 0/Not Present, this neighborhood would receive a code of .50 for this item. After all double-codes for all individual items within the neighborhood coding scheme were averaged to create a single code, items were composited into three composites (positive resources, safety, and

physical signs of order) by creating a sum score based on theory and previous research (McCoy et al., 2019). Ranges, means, standard deviations, frequencies and bivariate correlations among individual codes were examined to gain a descriptive understanding of children's proximal neighborhoods and to help inform further analyses.

Reliability and validity of the three coding composites were examined in several ways. First, internal consistency of the three composites was examined by calculating Cronbach's alpha for the individual items within each hypothesized composite. Next, construct validity of the coding composites was examined using several factor analyses. Previous literature (McCoy et al., 2019) demonstrated strong evidence for the factor structure of the signs of physical order composite, therefore, a confirmatory factor analysis (CFA) was performed for this composite in the present sample. However, there was no available previous literature that examined the factor structure of the positive resources or safety composites of the virtual neighborhood coding scheme. Therefore, to examine the factor structure of these composites, exploratory factor analyses (EFA) were performed for both of these composites in the present sample. Concurrent validity was assessed through examination of bivariate correlations among the three virtual neighborhood coding composites and the YWBI variables.

Research question 2. To examine if the assessment of children's proximal neighborhoods using the virtual neighborhood observation coding provides additional information over and above the assessment of neighborhoods using the census-travel level YWBI variables, we ran a fully unconditional two-level model. In this model, all child variables (i.e., virtual neighborhood coding variables, YWBI, and direct assessment of self-regulation) were at level-1 and the classroom was at level-2. Of particular interest was the calculation of the intraclass correlation coefficient (ICC). The ICC partitions the amount of variance that is

between the clusters (classrooms) versus within the cluster and provides information about the ability of the virtual neighborhood observation coding scheme to capture unique information about individual neighborhood characteristics for children who are in the same classroom. Separate models (three total) were run to calculate the ICCs for each construct within the virtual neighborhood observation coding scheme—physical signs of order, safety and resources. And, two separate models were run to calculate the ICCs for each of the YWBI variables—crime and housing indicators.

Research question 3. For research question 3, we were interested in the ability of the virtual neighborhood coding composites to predict growth in children's self-regulation skills over and above the census tract-level YWBI variables. To address this research question, we fit a two-level (Level 1= child, time; Level 2=classroom) model for change. First, to examine the trajectory of growth in children's self-regulation from the beginning to the end of the school year, a series of unconditional models for change were estimated. Next, a conditional multilevel model (MLM) was estimated where the growth parameters (intercept and slope) were regressed on the time-invariant variables (i.e., individual covariates). The census tract level YWBI indicators of crime levels and housing quality were entered as predictors of growth in self-regulation, controlling for all child demographics, family and home characteristics, and the classroom-level indicator of quality of teacher-child interactions. Finally, we added in the composites from the virtual neighborhood observation coding—physical signs of order, safety, and resources—as predictors of children's growth in self-regulation. This model was used as the basis for testing the full moderation model in the final research question below.

Research question 4. In our final research question, we were interested in examining if the relationship between resources and risk factors in children's proximal neighborhoods (as

measured by the virtual neighborhood coding) and children's growth in self-regulation differed depending upon the quality of children's interactions with their teacher in the classroom. To test the full moderation model, we estimated a conditional multi-level model with interactions to examine moderation by individual teacher-child interaction quality. Specifically, we interacted our virtual neighborhood coding composites with an aggregate (i.e., average over the year) of individual (i.e., inCLASS Teacher-Child Interactions domain) teacher-child interaction quality to predict our outcome, growth in self-regulation (three models in total were run). All predictors and moderators were grand-mean centered.

Results

Research Question 1: Exploring the Use of a New Coding System

Descriptive results of individual codes. Descriptive statistics were examined for all items within the coding scheme. First, for items that captured the physical order the neighborhood, we found that the quality of the majority of residences (51.4%) within neighborhoods were coded as being in good to excellent condition and only 2.5% of residences were coded as poor condition. With regards to the presence of well-manicured and cared for lawns, almost all (72.6%) residences in children's neighborhoods had lawns that were well-cared for. The majority of children's neighborhoods did not have abandoned cars or residences (87.5% and 79.1%, respectively). The condition of cars within neighborhoods was also good with over 50% of cars being coded as in good or excellent condition. In addition, most neighborhoods did not have abandoned pieces of large trash (61.7%) and most had light litter (74.1%). Lastly, most neighborhoods in the present sample did not have graffiti (95.6%) and the majority did not have bars on windows of any buildings (89.4%). See Table 2 for full descriptive characteristics. Bivariate correlations among the items within the hypothesized physical signs of order composite

varied and ranged from mild (r = .14 between abandoned trash and amount of litter) to strong (r = .68 between quality of residences and quality of lawns and r = .70 between condition of cars and abandoned cars; see Table 3 for all correlations within this composite).

The frequency of positive resources in neighborhoods varied, with the three most common resources across neighborhoods being religious institutions (64.4%), corner stores (48.9%) and K-12 schools (41.1%). On average, neighborhoods had 3.62 (SD= 2.65) positive resources (range: 0-11.50). Bivariate correlations among the positive resources were modest overall, with the strongest correlation (r = .49) being between financial institutions and sit-down restaurants (see Tables 4 and 5 for all correlation within this composite). Although there was limited variability in some of the positive resource codes (e.g., only 2.2% of neighborhoods had a library), all items were included in the factor analyses (described below) and in further analyses given that a lack of resources (or limited variability) is meaningful in describing children's neighborhoods and is hypothesized to negatively impact children's self-regulation development.

Lastly, the individual items with the safety composite showed limited variability. Almost all neighborhoods had a sidewalk on at least one side of the street (88.8%) and most neighborhoods had a crosswalk (71.9%). Fewer neighborhoods had speed bumps, a school zone sign or the presence of police (9.4%, 28.6%, 20.2%). Associations between items within the safety composite were generally low with the exception of presence of crosswalks and presence of sidewalks (r= .58; see Table 6).

Inter-rater reliability. All neighborhood maps were coded independently by two coders. All positive resource items are categorical (i.e., 1/Present; 0/Not Present) and were examined using an unweighted Kappa. The majority of codes for the positive resources indicated moderate agreement (Kappa > .41; Landis & Koch, 1977), with the exception of park, playground, athletic

complex, and hardware store (Kappa= .37, .35, .28 and .21, respectively; see Table 7). The physical signs of order composite contains codes that are both categorical and ordinal. Therefore, Kappa and ICC statistics were examined and reported accordingly. Kappas for the categorical codes (i.e., trash, graffiti, bars on windows, and abandoned cars and houses) demonstrated fair to moderate agreement (Kappa > .30) with the exception of abandoned cars, which exhibited poor agreement (Kappa= .16; See Table 8). ICCs for the ordinal codes within the signs of physical order composite ranged from low (ICC= .19 for amount of litter) to moderate (ICC= .62 for condition of houses). All codes within the safety composite are categorical (i.e., 1/Present; 0/Not Present) and were examined using an unweighted Kappa. Codes within the safety composite indicated mixed agreement. Codes for the presence of crosswalks and sidewalks demonstrated moderate to strong agreement (Kappa= .64 and .56, respectively), whereas codes for school zone sign, speed bump, and police presence indicated poor to fair agreement (Kappa= .24, .28 and .18, respectively; See Table 9).

Reliability and validity of coding composites. Given the strong association between residential and lawn quality and condition of cars and abandoned cars *and* findings from previous research, the covariances between these two sets of items were included in the CFA. Factor analysis (CFA) for the physical signs of order composite indicated adequate fit of the individual items to an overall latent factor reflecting the physical order of the neighborhood (χ^2 (25) = 103,929, p <.000; RMSEA = .099; CFI = .897; TLI = .851; SRMR = .048). These results are aligned with previous research examining the fit of these items to a latent factor (McCoy et al., 2019; χ^2 (33) = 123,868 p <.001; RMSEA = .097; CFI = .902; TLI = .867; SRMR = .054). Further, factor loadings for the individual items were good (range: .44- .80) with the exception of abandoned cars (.36). In addition, further analyses revealed moderate internal consistency of the

composite (Cronbach's alpha = .76; see Table 10). Given the fit of the overall model, the factor loadings, the alignment with previous research and the internal consistency of the composite, the physical signs of order composite was retained for use in further analyses.

For the positive resources composite, an EFA was examined. Results indicated four possible latent factors. Upon examination of fit indices and the scree plot along with a priori hypotheses and theory, two factors were retained for use in further analyses despite the suboptimal fit statistics of the model ($\chi^2(134) = 358{,}042$, $p < .000$; RMSEA = .063; CFI = .780; TLI = .719; SRMR = .056). The first factor contains items that are reflective of positive resources for public gathering spaces. This factor contains the following items (Geomin rotated loadings are provided in parentheses after each item; all are significant at 5% level): community center (.51), park (.37), playground (.18), athletic complex (.34), religious institution (.56), library (.19), police station (.30), and school (.33). Internal consistency of the eight items within the public gathering spaces factor was low (Cronbach's alpha = .54; see Table 10). The second retained factor contains items that are reflective of positive resources that help individuals meet basic and daily needs. This factor contains the following items (Geomin rotated loadings are provided in parentheses after each item; and, all are significant at 5% level): corner store (.20), grocery store (.42), financial institution (.67), sit-down restaurant (.67), salon/ barber shop (.44), pharmacy (.62), hardware store (.25), childcare/ daycare center (.46), adult learning center/college (.21), doctor's office (.56) and dentist's office (.68). Internal consistency of the eleven items within the basic and daily needs factor was moderate (Cronbach's alpha = .77; see Table 10). Lastly, the correlation between the two extracted factors was low ($r = .21$).

An EFA was also examined for the safety composite. Results indicated that one factor best fit the data. The fit of the model was strong ($\chi^2(5) = 17, 461$, $p < .005$; RMSEA = .088; CFI

= .919; TLI = .837; SRMR = .050), although individual factor loadings varied. This factor contains the following items (Geomin rotated loadings are provided in parentheses after each item and if significant, indicated by an asterisk): school zone sign (.12), presence of a crosswalk (.46*), presence of a sidewalk (1.25*), police presence (.083), and speed bump (.037). Internal consistency of the five items within the safety factor was poor (Cronbach's alpha = .42; see Table 10). Given the limited variability of items within this composite and the poor internal consistency, this composite was not retained for use in final models.

Lastly, associations between the retained neighborhood coding composites (i.e., physical signs of order, resources for basic and daily needs, and resources for public gathering) and the YWBI indicators were explored (see Table 11). With regards to resources that meet basic and daily needs, the more resources present in the area, the higher levels of crime in the surrounding census tract (r = -.28). Similarly, more resources available for public gathering was modestly associated with higher crime and lower housing quality in the surrounding census tract (r's= -.36 and -.33, respectively). With regards to physical signs of order, neighborhoods with more signs of order was modestly associated with better housing quality (r= .40) and lower crime (r= .21) in the surrounding census tract. Lastly, more signs of safety features was significantly and moderately associated with higher levels of crime (r= -.28) in the census tract.

Research Question 2: Comparison Between Coding Scheme and YWBI

Intraclass correlation coefficients were computed and compared to better understand the extent to which the neighborhood coding composites and the YWBI indicators captured unique information about the features of children's home neighborhoods for children who are in the same classroom. Results indicated that the virtual neighborhood coding captured slightly more unique information about the features of children's neighborhoods compared to the YWBI

indicators. Meaning that, compared the YWBI, there was more variability in the features of children's neighborhoods as captured by the coding (i.e., resources and physical signs of order) between children within a classroom (ICCs for basic and daily resources, public gathering spaces, physical signs of order, YWBI crime, YWBI housing are: .26, .30, .29, .33, .42, respectively; refer to Table 12).

Research Question 3: Direct Effects Between Neighborhood Indicators and Growth in Children's Self-Regulation

Descriptive statistics for all predictors and the outcome can be found in Tables 12 and 13. First, to ensure that there was significant variance in children's self-regulation at the beginning of the preschool year and with regards to their growth during the school year, an unconditional growth model using the three repeated measures of self-regulation (the outcome), including time (fall, winter, and spring of the school year) was examined. Results of this model indicated significant variance in children's self-regulation skills both at the beginning of the school year and in terms of their growth over the year (p's = .000).

Next, the direct effects model (i.e., direct effect of neighborhood features on growth in children's self-regulation) was examined. All covariates, predictors and the outcome were entered into the model simultaneously and the model demonstrated excellent fit ($\chi^2(14) =$ 22,460, $p >$. 05; RMSEA = .04; CFI = .978; TLI = .934; SRMR = .013). On average, at the beginning of the preschool year after controlling for all other covariates and accounting for nesting of children within classrooms, older children had significantly higher self-regulation skills (p = .000), boys had significantly lower self-regulation skills compared to girls ($p < .05$) and children who resided in census tracts with lower crime levels (i.e., YWBI Crime Indicator) had significantly higher self-regulation skills ($p < .05$). In terms of children's growth in self-

regulation over the preschool year, children with more resources for basic and daily needs had significantly more positive growth in self-regulation compared to children with less resources to meet basic and daily needs. There was no significant direct effect of the quality of children's individual interactions with their teacher on growth in their self-regulation skills.

Research Question 4: Combined Effect of Proximal Neighborhood Features and Children's Individual Interactions with Teachers on Self-Regulation Growth

Lastly, to examine if the relationship between proximal neighborhood features and children's growth in self-regulation was dependent upon the quality of individual interactions with teachers in the classroom, interaction terms between the three coding composites (i.e., resources to meet daily and basic needs, resources for gathering, and physical signs of order) and the quality of individual teacher-child interactions were entered into three separate models. One significant interaction between the number of resources to meet daily and basic needs and the quality of individual teacher-child interactions emerged (see Table 14 for detailed results of final significant interaction model). The model demonstrated excellent fit to the data ($\chi^2(15) =$ 23,838, $p >$. 05; RMSEA = .04; CFI = .978; TLI = .934; SRMR = .013). Results of the interaction indicated that for children who experienced lower-quality interactions with their teacher in the classroom *and* had few resources in their proximal neighborhoods to meet basic and daily needs, their self-regulation skills did not exhibit growth during the school year. However, for children who also had few resources in their proximal neighborhoods to meet basic and daily needs but experienced high-quality interactions with their teacher, their self-regulation skills grew significantly during the preschool year. Whereas, for children with many resources to meet basic and daily needs in their proximal neighborhoods, growth in their self-regulation skills did not vary depending on the quality of their interactions with their teacher in the classroom

(please refer to Figure 1 for a graph of the interaction effects).

Discussion

Children's early experiences across contexts (home, school, community) impacts their development and this is especially true for emergent brain systems that are responsible for important cognitive skills such as self-regulation (Blair & Raver, 2016). These experiences can hinder (e.g., food scarcity or poor and overcrowded housing) or bolster (e.g., warm, supportive relationships with adults or safe parks to play in and develop social-emotional skills with peers) this development (Cantor, Osher, Berg, Steyer, & Rose, 2018). Importantly, children's experiences across contexts are interactive (Lazarus & Launier, 1978); meaning that, if a child experiences feelings of stress walking through their neighborhood after school, then they might have a harder time sleeping at night which in turn affects their ability to control behaviors and feelings and pay attention in the classroom the next day. A deeper understanding of children's individual experiences within different contexts affords an opportunity to better understand where (e.g., in the classroom, in a community center, at home) and how (e.g., professional development aimed at improving the quality of teacher-child interactions, offering after-school programming, providing child care during parent classes) to leverage aspects of systems and resources already in place to support children's early development. This study advances the current literature through the use of novel assessment tools to better understand how features of children's proximal neighborhoods and individual experiences with their teacher combine to influence their self-regulation skills during a critical developmental period, preschool.

The results of the present study offer promise of the use of a virtual neighborhood coding scheme to measure features of children's proximal neighborhoods. The tool is freely available, can be completed from anywhere with an internet connection and requires little time. Results

provide some initial evidence of the combined importance of *both* the resources available in preschool children's immediate neighborhoods and their interactions with their teacher in the classroom for the development of their self-regulation skills. Importantly, this combined effect was only present for the proximal neighborhood resources and not for the census tract-level neighborhood features (i.e., Youth Well-Being Index crime levels and housing quality) highlighting the potential significance of the proximity of resources to a family's home and in particular those resources that help to meet basic and daily needs.

Use of a Novel Virtual Coding Scheme: Lessons Learned and Next Steps

Overall, the implementation of the virtual coding scheme provided more specificity in the features of children's neighborhoods than is typically available. In doing so, we were better able to understand the composition of neighborhoods that children interface with every day. However, there were challenges in obtaining adequate variability within individual codes and achieving strong reliability and validity of the hypothesized coding composites. These challenges and suggestions for future research are discussed in more detail below.

Neighborhood composition. Children's neighborhoods in the present sample were somewhat different from the neighborhoods being used in the larger validation sample of the tool (McCoy et al., 2019). First, we found that children's neighborhoods in the current study were more suburban and less densely populated compared to much of the sample of neighborhoods in current work by McCoy and colleagues (2019). Therefore, neighborhoods in the current study tended to be more residential and have fewer mixed-use spaces (i.e., residential and commercial/ business properties) which resulted in fewer available resources (i.e., grocery stores, libraries, parks) per proximal neighborhood. However, the lack of resources present in many neighborhoods is meaningful given that the majority of the present sample included children

from low-income families and it can be a challenge (e.g., caregivers working several jobs) for caregivers to regularly and successfully (e.g., reliable transportation or easy bus routes) access resources to meet every day needs. Further, the ICCs between the neighborhood coding composites and the YWBI census tract variables indicated that even for children in the same classroom (and therefore residing in generally the same larger neighborhoods), the neighborhood coding provided somewhat more unique information about children's proximal neighborhoods close to their homes—and, this level of specificity was important for children's self-regulation development. However, this research is burgeoning and the results of the present study highlight the importance of future work to partner closely with communities to better understand how families access resources. In other words, is it burdensome to not have many resources available in one's proximal neighborhood? Or, perhaps given the context (i.e., more suburban or rural), most families are able to easily drive further away to access resources.

With regards to safety features present in the neighborhoods, the interpretation challenges were twofold. First, and similar to the resources, there was very limited variability in certain individual codes—specifically, almost none of the neighborhoods had speed bumps, school zone signs or police presence, while almost all of the neighborhoods had a sidewalk present on at least one side of the road and at least one crosswalk. The skewness of the data for this composite limited our ability to combine the individual codes into a meaningful composite. Additionally and importantly, without input from community members, it is difficult to interpret the safety codes within the context of a particular community. In other words, in some neighborhoods, a police presence might be reassuring while in other communities, the presence of police is disconcerting and scary. Due to the limited variability and interpretation concerns, the safety composite was not used in further analyses in the present paper. However, it is important that

future work better examine how individual communities quantify feelings of safety given the well-established link between feelings of safety (or lack thereof), stress, and blood cortisol levels, all of which in turn impact the cognitive systems responsible for self-regulation development (Blair et al., 2011).

Challenges in obtaining coder reliability. For this particular coding scheme, examination of IRR was critical because previous work (McCoy et al., 2019) found that coders' experiences in their own neighborhoods influenced the ways in which they coded neighborhoods using the virtual tool. For example, if a coder grew up in a neighborhood where the sidewalks were re-paved every spring, their impression of sidewalk quality likely differ from another coder who grew up in a neighborhood where public funds were not allocated to regular upkeep of sidewalks. We found that regular and consistent coder meetings were important to maintain calibration within the coding team. We also double-coded all neighborhoods to help ensure reliability. With these steps in place, we were able to achieve good IRR across most items within the coding scheme, however, the subjective codes (e.g., rating residential quality on a scale from poor- excellent) remained more challenging to achieve perfect agreement. The inter-coder differences also present an opportunity for future work to focus specifically on the ways in which previous experiences influence perceptions of neighborhoods and what the implications are for how interpretations are made about neighborhood quality.

Links Between Neighborhood Resources and Physical Signs of Order and Children's Self-Regulation Skills

Consistent with our hypothesis, we found that resources in the blocks immediately surrounding children's homes that help to meet basic and daily needs (i.e., grocery stores, doctor's offices, corner stores) were significantly associated with growth in children's self-

regulation skills over and above more typically used census-tract level indicators of neighborhoods. While mechanisms or pathways for this effect could not be examined in the present study, it can be hypothesized that perhaps for parents or caregivers having easy access to resources such as food or a pharmacy for medication or a doctor's office when sick lessens daily stressors. And, previous research (Blair et al., 2011) indicates that when caregivers experience higher levels of stress, it is more difficult to manage children's behaviors and maintain routines—behaviors that in turn impact children's own self-regulation development. It is also possible that there is something about the families themselves (e.g., income, maternal education, household climate) that fosters self-regulation development in children and these families are disproportionately overrepresented in neighborhoods with more resources. Or, perhaps having access to fresh food, medicine from a pharmacy or a doctor's office improves physical health (e.g., activity level, obesity, lower blood cortisol) which promotes the development of the cognitive systems needed for self-regulation (Diamond & Lee, 2011). Future research should examine these pathways more specifically.

Surprisingly, we did not find significant direct effects for the other two out of the three retained composites of the neighborhood coding scheme (i.e., resources for public gathering or physical signs of order). There are two potential reasons why these effects may not have emerged in the present sample—one is due to the lack of variability within the items in these composites and the other is due to the specific sample and neighborhoods in the present study. The individual items within the resources for public gathering spaces demonstrated limited variability, meaning that, many neighborhoods simply did not have parks or playgrounds. This limited variability makes it difficult to detect effects on children's self-regulation development because all children in the sample received similar codes. It could also be that given the types of

neighborhoods in the present sample, most children may have access to front or back yards to play in and therefore the lack of presence of parks or playgrounds is not as important for self-regulation development as it might be for children who resides in a high-rise apartment complex in an urban center. Similarly, for the physical signs of order composite, there was limited variability in individual codes. Most neighborhoods had sidewalks and crosswalks in good condition and overall the housing quality was good. And, having sidewalks or crosswalks might be less important for children who primarily drive with their caregivers to various locations versus children who rely on safe sidewalks every day to walk to a subway or bus stop.

Overall, much more work is needed to explore the mechanisms through which the proximity of resources in family's neighborhood influences the development of young children's self-regulation skills. This line of research has implications that include things such as how developmental psychologists might work together with city planners to carefully craft and include resources in low-income communities that might help to alleviate daily stressors related to meeting basic needs and allow caregivers to have more time to spend engaged with children during this critical time period. However, we know that to most effectively bolster young children's self-regulation, it is critical to look not only at how to support children in one context (such as their neighborhood) but to carefully consider how different contexts (e.g., neighborhoods and classrooms) interact to influence children's self-regulation development.

The Combined Effect on Self-Regulation Skills of Resources for Basic and Daily Needs and Children's Individual Interactions with Teachers

We found that the quality of a child's interactions with their teacher in the classroom may be especially important for children who reside in neighborhoods with limited resources to meet basic and daily needs. The experience of having warm and supportive interactions with a teacher

in the classroom may reduce feelings of stress during the school day for young children who are more likely to experience stress outside of school, which in turn provides them with the opportunity to practice and develop self-regulation skills. This perhaps suggests that while the experience of residing in an under-resourced neighborhood might negatively impact self-regulation skills, that high quality individual experiences in the classroom might be protective and have the potential to support children's growth in this area.

While the other coding composites (i.e., resources for gathering and physical signs of order) did not directly affect or significantly interact with children's individual experiences with their teacher to influence their self-regulation skills, it does not necessarily suggest that these features of neighborhoods are not important for children's self-regulation development but rather perhaps from a perspective of hierarchy of needs (i.e., Maslow, 1943), meeting basic needs is most important for the cognitive systems responsible for fundamental self-regulation development. And, as children grow older and the use of self-regulation skills becomes more complex (i.e., appropriately inhibiting during conversations with peers or refraining from calling out in class), the importance of other resources such as community centers or parks that offer multiple opportunities for social engagement might become much more important. Future research should aim to follow children longitudinally to better how neighborhood features interact with children's individual experiences in the classroom to influence the development of self-regulation skills.

Limitations

There are several limitations to be noted in the present study. First with regards to the neighborhood coding tool, the neighborhoods in the present study were more suburban and rural compared to the larger sample that the tool is currently being validated with. Therefore, for the

items with poor reliability in the present study, it is difficult to assert whether there was something about the coders in our study or if these items do not function well in more suburban/rural neighborhoods. In addition, our factor analyses were helpful in guiding initial construction of the theorized composites in this study but a larger sample size across diverse neighborhoods (i.e., urban/suburban; wealthy/poor) would offer further evidence of the psychometric properties of the tool. The sample in the present study was also relatively small for what is needed to examine the validity of a new measurement tool and the sample was limited to one region in one Southeastern US state.

Future Research and Next Steps

Future research that examines how the features of children's proximal neighborhoods impact early development should seek to broaden the sample to include diverse geographic areas (i.e., across the U.S. and internationally) and a range of neighborhood types (i.e., rural, suburban, and large, dense urban centers). It is possible that the coding scheme as it currently is constructed is better suited to assessing features of urban neighborhoods and children residing in more suburban or rural neighborhoods may face challenges in their communities that are not presently captured. For example, obtaining information regarding car ownership may offer insight into how difficult or easy it is for a family to access resources. In addition, community-research partnerships wherein community members are included in both measure construction and interpretation is critical in understanding where and how to support families and communities, especially communities that are traditionally underserved and marginalized.

Conclusion

The current study offers some initial evidence that suggests that it might be important to capture features of children's proximal neighborhoods for better understanding individual early

development. Many of young children's everyday activities take place within the area immediately surrounding their homes; for example, they play outside with friends in the neighborhood, or their caregivers might take them to the local library or park. And, caregivers also need to access resources on a regular basis such as the grocery store or pharmacy or doctor—tasks that are far easier and less stressful if these resources are easily accessible. This approach to capturing proximal resources of neighborhoods offers more insight and nuance about the everyday experiences of children and families than is available through larger data such as census tract-level neighborhood data. In addition, the present study highlights the importance of capturing positive features of neighborhoods in addition to risk factors. This approach to understanding neighborhoods has important implications for how future work might leverage features already present in neighborhoods (e.g., an afterschool program in a community center for young children or parenting classes at a local library) and help to inform future development of neighborhoods (e.g., amount of green space or a fresh food store) to best support young children's development early on so they can thrive in school and life.

In addition to neighborhoods, children spend a large amount of their time in their classroom with their teacher. We found that the experience of a warm, and supportive teacher can be protective for children who reside in neighborhoods with few basic resources. Overall, these findings have implications for policymakers, researchers and stakeholders to carefully consider all aspects of children's early lives (e.g., where do children live and what is or is not available nearby, how safe are children's schools and do they feel supported by teachers) to best understand how to support children, families and schools to be ready to thrive in school and beyond.

References

Ackerman, B. P., Izard, C. E., Schoff, K., Youngstrom, E. A., & Kogos, J. (1999). Contextual risk, caregiver emotionality, and the problem behaviors of six- and seven-year-old children from economically disadvantaged families. *Child Development, 70*(6), 1415-1427.

Baker, J. A., Grant, S., & Morlock, L. (2008). The teacher-student relationship as a developmental context for children with internalizing or externalizing behavior problems. *School Psychology Quarterly, 23*, 3–15

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. (2015). School readiness and self-regulation: A developmental psychobiological approach. *Annual Review of Psychology*, *66*, 711-731.

Blair, C., Raver, C. C., Granger, D., Mills-Koonce, R., Hibel, L., & Family Life Project Key Investigators. (2011). Allostasis and allostatic load in the context of poverty in early childhood. *Development and Psychopathology, 23*(3), 845-857.

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

Bronfenbrenner, U. (1988). Interacting systems in human development. Research paradigms: Present and future. *Persons in Context: Developmental Processes*, 25-49.

Bronfenbrenner, U., & Ceci, S. J. (1994). Nature-nurture reconceptualized in developmental perspective: A bioecological model. *Psychological review, 101*(4), 568.

Bronson, M. B. (2000). Recognizing and supporting the development of self-regulation in young children. *Young Children, 55*(2), 32-37.

Brown, E. D., Ackerman, B. P., & Moore, C. A. (2013). Family adversity and inhibitory control for economically disadvantaged children: Preschool relations and associations with school readiness. *Journal of Family Psychology, 27*(3), 443.

Buyse, E., Verschueren, K., & Doumen, S. (2011). Preschoolers' attachment to mother and risk for adjustment problems in kindergarten: Can teachers make a difference? *Social Development, 20*(1), 33-50.

Cadima, J., Verschueren, K., Leal, T., & Guedes, C. (2016). Classroom interactions, dyadic teacher–child relationships, and self–regulation in socially disadvantaged young children. *Journal of Abnormal Child Psychology, 44*(1), 7-17.

Campbell, S. B., Denham, S. A., Howarth, G. Z., Jones, S. M., Whittaker, J. V., Williford, A. P., ... & Darling-Churchill, K. (2016). Commentary on the review of measures of early childhood social and emotional development: Conceptualization, critique, and recommendations. *Journal of Applied Developmental Psychology*, *45*, 19-41. doi: 10.1016/j.appdev.2016.01.008

Cantor, P., Osher, D., Berg, J., Steyer, L., & Rose, T. (2018). Malleability, plasticity, and individuality: How children learn and develop in context. *Applied Developmental Science*, 1-31.

Caughy, M. O. B., Leonard, T., Beron, K., & Murdoch, J. (2013). Defining neighborhood boundaries in studies of spatial dependence in child behavior problems. *International Journal of Health Geographics*, *12*, 24.

Coulton, C. J., Richter, F., Kim, S. J., Fischer, R., & Cho, Y. (2016). *Leveraging Integrated Data Systems to Examine the Effect of Housing and Neighborhood Conditions on Kindergarten Readiness.* Center on Urban Poverty and Community Development, Jack,

Joseph and Morton Mandel School of Applied Social Sciences, Case Western Reserve University.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297-334.

Cutrona, C. E., Wallace, G., & Wesner, K. A. (2006). Neighborhood characteristics and depression: An examination of stress processes. *Current Directions in Psychological Science*, *15*(4), 188-192.

Daily, S., Burkhauser, M. & Halle, T. (2010). A review of school readiness practices in the States: Early Learning Guidelines and Assessments. Early Childhood Highlights (Child Trends), 1(3). Retrieved from: http://www.childtrends.org/wpcontent/uploads/2013/05/2010-14-SchoolReadinessStates.pdf

Davis, M., Bilms, J., & Suveg, C. (2017). In sync and in control: A meta-analysis of parent–child positive behavioral synchrony and youth self-regulation. *Family process*, *56*(4), 962-980.

Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4 to 12 years old. *Science*, *333*, 959-964.

Downer, J. T., Booren, L. M., Lima, O. K., Luckner, A. E., & Pianta, R. C. (2010). The Individualized Classroom Assessment Scoring System (inCLASS): Preliminary reliability and validity of a system for observing preschoolers' competence in classroom interactions. *Early Childhood Research Quarterly, 25*, 1-16. doi: 10.1016/j.ecresq.2009.08.004.

Duncan, R. J., Schmitt, S. A., Burke, M., & McClelland, M. M. (2018). Combining a kindergarten readiness summer program with a self-regulation intervention improves school readiness. *Early Childhood Research Quarterly*, *42*, 291-300.

Eisenberg, N., Valiente, C., & Eggum, N. D. (2010). Self-regulation and school readiness. *Early Education and Development, 21*, 681-698. doi: 10.1080/10409289.2010.497451.

Evans, G. W. (2003). A multimethodological analysis of cumulative risk and allostatic load among rural children. *Developmental Psychology, 39*, 924- 933.

Evans, G. W., & Kim, P. (2013). Childhood poverty, chronic stress, self-regulation, and coping. *Child Development Perspectives*, *7*(1), 43-48.

Fassaert, T., De Wit, M., Tuinebreijer, W., Wouters, H., Verhoeff, A., Beekman, A., & Dekker, J. (2009). Psychometric properties of an interviewer-administered version of the Kessler Psychological Distress scale (K10) among Dutch, Moroccan and Turkish respondents. *International Journal of Methods in Psychiatric Research, 18*, 159-168. doi: 10.1002/mpr.288.

Flouri, E., Midouhas, E., & Joshi, H. (2014). The role of urban neighbourhood green space in children's emotional and behavioural resilience. *Journal of Environmental Psychology, 40,* 179-186. doi: 10.1016/j.jenvp.2014.06.007.

Fox, L., Dunlap, G., Hemmeter, M. L., Joseph, G. E., & Strain, P. S. (2003). The Teaching Pyramid: A model for supporting social competence and preventing challenging behavior in young children. *Young Children, 58*, 48-52.

Franco, L., Pottick, K., & Huang, C. C. (2010). Early parenthood in a community context: Neighborhood conditions, race–ethnicity, and parenting stress. *Journal of Community Psychology*, *38*(5), 574-590.

Garmezy, N., & Rutter, M. (1983). *Stress, coping, and development*. New York, NY: McGraw-Hill.

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and

tutorial. *Tutorials in Quantitative Methods for Psychology, 8*, 23-34.

Hamre, B., Hatfield, B., Pianta, R., & Jamil, F. (2014). Evidence for general and domain-specific elements of teacher–child interactions: Associations with preschool children's development. *Child Development, 85*, 1257-1274. doi: 10.1111/cdev.12184.

Hamre, B. K., & Pianta, R. C. (2001). Early teacher–child relationships and the trajectory of children's school outcomes through eighth grade. *Child Development, 72*, 625-638. doi: 10.1111/1467-8624.00301.

Hamre, B. K., & Pianta, R. C. (2005). Can instructional and emotional support in the first-grade classroom make a difference for children at risk of school failure? *Child Development, 76*, 949-967. doi: 10.1111/j.1467-8624.2005.00889.x.

Heberle, A. E., Thomas, Y. M., Wagmiller, R. L., Briggs-Gowan, M. J., & Carter, A. S. (2014). The impact of neighborhood, family, and individual risk factors on toddlers' disruptive behavior. *Child Development, 85*, 2046-2061. doi: 10.1111/cdev.12251.

Jones, A. P., Coombes, E. G., Griffin, S. J., & van Sluijs, E. M. (2009). Environmental supportiveness for physical activity in English schoolchildren: A study using Global Positioning Systems. *International Journal of Behavioral Nutrition and Physical Activity, 6*, 42.

Kessler, R. C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Normand, S. L., ... Zaslavsky, A. M. (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine, 32*, 959-976. doi: 10.1017/S0033291702006074.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159-174.

Lazarus, R. S., & Launier, R. (1978). Stress-related transactions between person and environment. In *Perspectives in Interactional Psychology* (pp. 287-327). Springer, Boston, MA.

Maslow, A. H. (1943). A theory of motivation. *Psychological review*, *50*(4), 370-396.

Matheny, A. P., Wachs, T. D., Ludwig, J. L., & Phillips, K. (1995). Bringing order out of chaos: Psychometric characteristics of the confusion, hubbub, and order scale. *Journal of Applied Developmental Psychology, 16*, 429-444. doi: 10.1016/0193-3973(95)90028-4.

McClelland, M. & Cameron, C. (2012). Self-regulation in early childhood: Improving conceptual clarity and developing ecologically valid measures. *Child Development Perspectives, 6*, 136-142. doi:10.1111/j.1750-8606.2011.00191.x

McCoy, D. C., Connors, M. C., Morris, P. A., Yoshikawa, H., & Friedman-Krauss, A. H. (2015). Neighborhood economic disadvantage and children's cognitive and social-emotional development: Exploring Head Start classroom quality as a mediating mechanism. *Early Childhood Research Quarterly, 32*, 150-159. doi: 10.1016/j.ecresq.2015.04.003.

McCoy, Sabol, Hanno, & Odgers (2019). Virtual school neighborhood observation: A digital tool for estimating and understanding the impact of schools and their neighborhoods. *In Progress.*

McEwen, B. S. (2013). The brain on stress: Toward an integrative approach to brain, body, and behavior. *Perspectives on Psychological Science*, *8*(6), 673-675.

Minh, A., Muhajarine, N., Janus, M., Brownell, M., & Guhn, M. (2017). A review of neighborhood effects and early child development: How, where, and for whom, do neighborhoods matter? *Health & Place, 46*, 155-174. doi: 10.1016/j.healthplace.2017.04.012.

Muthén, L. K. (2018). Mplus Users Guide. Los Angeles, CA: Muthén & Muthén; 2010. *Computer software and manual*.

National Education Goals Panel (1995). Reconsidering children's early development and learning: Toward common views and vocabulary. Retrieved from: http://govinfo.library.unt.edu/negp/reports/child-ea.htm

O'Connor, E., & McCartney, K. (2007). Examining teacher–child relationships and achievement as part of an ecological model of development. *American Educational Research Journal, 44*, 340-369. doi: 10.3102/0002831207302172.

Odgers, C. L., Caspi, A., Bates, C. J., Sampson, R. J., & Moffitt, T. E. (2012). Systematic social observation of children's neighborhoods using Google Street View: A reliable and cost-effective method. *Journal of Child Psychology and Psychiatry, 53*, 1009-1017. doi: 10.1111/j.1469-7610.2012.02565.x.

Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher, 38*, 109-119. doi: 10.3102/0013189X09332374.

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System™: Manual K-3*. Paul H Brookes Publishing.

Pianta, R. C., & Stuhlman, M. W. (2004). Teacher-child relationships and children's success in the first years of school. *School Psychology Review, 33*, 444.

Ponitz, C. E. C., McClelland, M. M., Jewkes, A. M., Connor, C. M., Farris, C. L., & Morrison, F. J. (2008). Touch your toes! Developing a direct measure of behavioral regulation in early childhood. *Early Childhood Research Quarterly, 23*, 141-158. doi: 10.1016/j.ecresq.2007.01.004.

Ratcliffe, M., Burd, C., Holder, K., & Fields, A. (2016). Defining rural at the US Census Bureau. *American Community Survey and Geography Brief*, 1-8.

Raudenbush, S. W., & Bryk, A. S. (2002). Hierarchical linear models: Applications and data analysis methods (Vol. 1). Sage.

Raudenbush, S. W., & Sampson, R. J. (1999). Ecometrics: Toward a science of assessing ecological settings, with application to the systematic social observation of neighborhoods. *Sociological Methodology, 29*(1), 1-41.

Rimm-Kaufman, S. E., Curby, T. W., Grimm, K. J., Nathanson, L., & Brock, L. L. (2009). The contribution of children's self-regulation and classroom quality to children's adaptive behaviors in the kindergarten classroom. *Developmental Psychology, 45*(4), 958.

Rimm-Kaufman, S. E., La Paro, K. M., Downer, J. T., & Pianta, R. C. (2005). The contribution of classroom setting and quality of instruction to children's behavior in kindergarten classrooms. *The Elementary School Journal, 105*(4), 377-394.

Sabol, T. J., & Pianta, R. C. (2012). Recent trends in research on teacher–child relationships. *Attachment & Human Development, 14*(3), 213-231.

Scott, R. (2011). The role of public libraries in community building. *Public Library Quarterly*, *30*(3), 191-227.

Stetsenko, A., & Vianna, E. (2009). Bridging developmental theory and educational practice: Lessons from the Vygotskian Project. In O. A. Barbarin & B. Hanna Wasik (Eds.), Handbook of child development and early education: Research to practice (pp. 38–54). New York, NY: Guilford Press.

Taylor, A. F., Kuo, F. E., & Sullivan, W. C. (2002). Views of nature and self-discipline: Evidence from inner city children. *Journal of Environmental Psychology, 22*(1-2), 49-63.

U.S. Census Bureau (2015). Census tracts and block numbering areas. In Geographic Areas Reference Manual (Chapter 10). Retrieved from https://www.census.gov/geo/reference/garm.html

U.S. Department of Health and Human Services (2015). The Head Start child development and early learning framework (Office of Head Start, Administration for Children and Families). Arlington, VA: Head Start Resource Center.

U.S. Department of Health and Human Services (2016). Head Start Performance Standards (Office of Head Start, Administration for Children and Families, 45 CFR Chapter XIII). Arlington, VA.

Vaden-Kiernan, M., D'elio, M. A., O'brien, R. W., Tarullo, L. B., Zill, N., & Hubbell-McKey, R. (2010). Neighborhoods as a developmental context: A multilevel analysis of neighborhood effects on Head Start families and children. *American Journal of Community Psychology, 45(*1-2), 49-67.

Vernon-Feagans, L., & Cox, M. (2013). Poverty, rurality, parenting, and risk: An introduction. *Monographs of the Society for Research in Child Development*, *78*(5), 1-23.

Virginia Department of Health. (2017). [Graph illustration of the Youth Well Being Index by County/ Independent City]. Virginia Health Opportunity Index. Retrieved from https://www.vdh.virginia.gov/omhhe/hoi/youth-well-being-index

Williford, A., Whittaker Vick, J., Vitiello, V. & Downer, J. (2013). Children's engagement within the preschool classroom and their development of self-regulation. *Early Education and Development, 24*, 162-187. doi:10.1080/10409289.2011.628270

Tables and Figures

Table 1
Participant Characteristics

	%	Mean	SD	Range
Child Race				
Black	59.1			
White	32.1			
Hispanic	8.8			
Families with annual income < $20, 000	34.2			
Child Age in the Fall (months)		52.51	3.72	40-67
Child Sex (male)	49.9			
Teacher Sex (female)	98.0			
Teacher Age (years)		44.59	10.43	23-63
Teacher Race				
White	69.8			
Black	22.6			
Hispanic	3.8			

Note. Data were collected on 380 preschool children from 51 classrooms in one southern state.

Table 2
Descriptive Statistics of Neighborhood Coding Items

Item	% (present)	Mean	SD
Park	24.6		
Playground	23.4		
Athletic Complex	22.1		
Community Center	16.5		
Religious Institution	64.4		
Grocery Store	26.7		
Corner Store	48.9		
Financial Institution	23.1		
Sit-down	33.9		
Salon/ Barbershop	33.3		
Pharmacy	17.8		
Hardware Store	9.1		
Library	5.3		
Police/ Fire Station	14.4		
Daycare Center	17.5		
K-12 School	41.1		
College/ Adult Learning Center	3.2		
Doctor's Office	22.4		
Dentist's Office	13.7		
Residential Quality		.55	.18
Lawn Quality		.67	.21
Abandoned Residences	8.4		
Trash	19.0		
Litter		.84	.29
Graffiti	3.1		
Bars on Windows	10.6		
Abandoned Cars	7.2		
Condition of Cars		.52	.18
School Zone Sign	28.6		
Speed Bump	9.4		
Crosswalk	71.9		
Sidewalk	88.8		
Police Presence	20.2		

Note. Residential Quality, Lawn Quality, and Condition of Cars were all re-coded to be on a 0-1 scale with 0= poor and 1= excellent. Litter was re-coded to be on a 0-1 scale with 0= heavy and 1= light. All other items were coded 0 (not present) or 1 (present).

Table 3
Bivariate Correlations Among Individual Items Within the Physical Signs of Order Composite

	1	2	3	4	5	6	7	8	9
1. Litter	1	.14*	.32**	.45**	.34**	.25**	.17**	.36**	.24**
2. Trash		1	.24**	.34**	.24**	.37**	.11	.23**	.34**
3. Quality of Lawns			1	.68**	.37**	.34**	.23**	.40**	.46**
4. Condition of Houses				1	.35**	.38**	.24**	.56**	.49**
5. Graffiti					1	.38**	.38**	.33**	.36**
6. Bars on Windows						1	.15**	.31**	.41**
7. Abandoned Cars							1	.70**	.26**
8. Condition of Cars								1	.33**
9. Abandoned Houses									1

Note. *p< .05, two-tailed. **p< .01, two-tailed.

Table 4
Bivariate Correlations Among Individual Items Within the Gathering Places Composite

	1	2	3	4	5	6	7	8
1. Park	1	.38**	.20**	.17**	.22**	.16**	-.01	.06
2. Playground		1	.19**	.04	-.002	.12*	-.02	-.01
3. Athletic Center			1	.19**	.14*	.34**	.04	.11
4. Community Center				1	.24**	.20**	.23**	.10
5. Religious Institution					1	.02	.23**	.30**
6. Library						1	.10	.03
7. School							1	.08
8. Police/ Fire								1

Note. *p< .05, two-tailed. **p< .01, two-tailed.

Table 5
Bivariate Correlations Among Individual Items Within the Basic and Daily Needs Resource Composite

	1	2	3	4	5	6	7	8	9	10	11
1. Grocery Store	1	.14**	.25**	.28**	.31**	.43**	.11	.10	.11	.16**	.21**
2. Corner Store		1	.20**	.34**	.26**	.23**	.13*	.13*	-.01	.12*	.05
3. Financial Institution			1	.49**	.33**	.38**	.19**	.30**	.17**	.34**	.41**
4. Sit-down Restaurant				1	.43**	.45**	.22**	.30**	.21**	.34**	.38**
5. Salon/ Barber Shop					1	.29**	.11*	.33**	.08	.20**	.26**
6. Pharmacy						1	.13*	.30**	.02	.36**	.39**
7. Hardware Store							1	.04	.01	.18**	.20**
8. Child Care Center								1	.09	.33**	.37**
9. College/ Adult Learning Center									1	.07	.19**
10. Doctor's Office										1	.50**
11. Dentist's Office											1

Note. *p< .05, two-tailed. **p< .01, two-tailed.

Table 6
Bivariate Correlations Among Individual Items Within the Safety Composite

	1	2	3	4	5
1. School Zone Sign	1	.07	.08	.16**	-.15**
2. Speed Bump		1	.01	.05	.14*
3. Crosswalk			1	.58**	.03
4. Sidewalk				1	.10
5. Police Presence					1

Note. *p< .05, two-tailed. **p< .01, two-tailed.

Table 7
Inter-rater Reliability Statistics for the Positive Resource Items of the Neighborhood Coding

Item	Kappa
Corner Store	.57
Library	.53
Park	.37
Religious Institution	.66
Playground	.35
Athletic Complex	.28
Grocery Store	.60
Sit-Down Restaurant	.54
Financial Institution	.63
Barber Shop/ Salon	.55
Pharmacy	.78
Community Center	.46
Hardware Store	.21
Library	.53
Police/Fire Station	.63
Daycare Center	.50
K-12 School	.53
College	.61
Doctor's Office	.60
Dentist's Office	.60

Note. All items are coded as 0/1: Present or Not Present.

Table 8
Inter-rater Reliability Statistics for Physical Signs of Order Items of the Neighborhood Coding

Item	Coding	ICC	Kappa
Litter	Light, Moderate, Heavy	.19	N/A
Trash	Present, Not Present	N/A	.47
Quality of Lawns	Poor, Fair, Good, Excellent	.47	N/A
Condition of Houses	Poor, Fair, Good, Excellent	.62	N/A
Graffiti	Present, Not Present	N/A	.33
Painted Over Graffiti	Present, Not Present	N/A*	N/A*
Bars on Windows	Present, Not Present	N/A	.49
Abandoned Cars	Present, Not Present	N/A	.16
Condition of Cars	Poor, Fair, Good, Excellent	.41	N/A
Abandoned Houses	Present, Not Present	N/A	.30

Note. All items were re-scaled to be on a 0-1 scale with higher scores representing more signs of physical order. *Not present in any neighborhoods and not included in calculation of sum score.

Table 9
Inter-rater Reliability Statistics for Safety Items of the Neighborhood Coding

Item	Kappa
School Zone Sign	.24
Speed Bump	.28
Crosswalk	.64
Police	.18
Sidewalk	.56

Note. All items are coded as 0/1: Present or Not Present.

Table 10
Internal Consistency of Neighborhood Coding Composites

Neighborhood Coding Composite	Cronbach's Alpha
Resources for Public Gathering Spaces	.54
Resources for Basic and Daily Needs	.77
Physical Signs of Order	.76
Safety	.42

Table 11
Bivariate Correlations Among Final Coding Composites and the YWBI Indicators

	1	2	3	4	5
1. Resources for Gathering	1	.25**	-.20**	-.36**	-.33**
2. Resources for Basic and Daily Needs		1	.03	-.28**	.03
3. Physical Signs of Order			1	.21**	.40**
4. YWBI Crime				1	.45**
5. YWBI Housing Quality					1

Note. ** $p < .01$, two-tailed. All virtual neighborhood coding composites are sum scores and higher scores indicate more resources, more physical signs of order or the presence of more safety features. Higher scores on the YWBI indicate less crime and higher housing quality.

Table 12
Descriptive Statistics for Neighborhood Coding Composites and YWBI Variables

	Min.	*Max.*	*Mean*	*SD*	*ICC*
Virtual Neighborhood Coding Composites					
Resources for Gathering	0	5.50	1.61	1.26	.30
Resources for Basic and Daily Needs	0	8.00	2.01	2.04	.26
Physical Signs of Order	0	8.83	6.90	1.47	.29
Safety Features	0	4.50	1.85	.88	.20
Youth Well-Being Index (YWBI) Indicators					
Housing Quality Index	.11	.82	.53	.16	.42
Crime Index	.25	.81	.45	.10	.33

Note. Intraclass correlation coefficients (ICC) were clustered at the classroom-level. All virtual neighborhood coding composites are sum scores and higher scores indicate more resources, more physical signs of order or the presence of more safety features. Higher scores on the YWBI indicate less crime and higher housing quality.

Table 13
Descriptive Statistics of HTKS, and inCLASS in the Fall, Winter and Spring

Measure	Mean	SD	Range
HTKS Fall	18.03	22.02	0-87
HTKS Winter	31.71	27.16	0-94
HTKS Spring	40.40	29.45	0-94
InCLASS Teacher-Child Intx Fall	2.06	.67	1-4.42
InCLASS Teacher-Child Intx Winter	2.21	.78	1-5.67
InCLASS Teacher-Child Intx Spring	2.02	.85	1-5.83

Note. YWBI indicators are scaled on a 0-1 scale with higher scores reflecting less crime and better housing quality. The HTKS scores at each timepoint (fall, winter, spring) include the practice items.

Table 14

Standardized Associations Between Neighborhood Features, Interaction Between inCLASS and Resources for Daily and Basic Needs and Children's Self-Regulation Skills

	Self-Regulation in the Fall of Preschool (Intercept)		Growth in Self-Regulation (Slope)	
	Est.	*SE*	Est.	*SE*
Covariates				
Gender (boy)	-6.18**	2.32	-1.80	1.22
Race (Black)	-.94	3.13	-3.31	1.74
Age (months)	1.09***	.27	.16	.16
Income-to-needs Ratio	.44	1.32	-.92	.66
Maternal Education	-.03	.86	.71	.48
Residential Chaos	.90	1.96	-.57	.63
Census Tract Neighborhood Features				
YWBI Low Crime	2.54*	1.26	9.21***	1.97
YWBI Housing Quality	1.30	1.45	4.43*	1.97
Proximal Neighborhood Coding				
Resources for Basic and Daily Needs	.92	1.35	1.70**	.61
Resources for Public Gathering	.10	1.31	1.49.	.78
Physical Signs of Order	1.63	1.31	-.15	.69
Classroom Observations				
InCLASS T-Ch Interactions	1.71	1.11	.48	.77
CLASS Emotional Support	1.31	1.19	-2.64***	.76
Interaction				
InCLASS*Basic and Daily Needs	2.34	1.32	-1.55*	.65

Note. *** $p = .000$, ** $p < .01$, * $p < .05$. All variables were entered into the model simultaneously. All predictors and the interactions were z-scored and standardized for ease of interpretation. The outcome (self-regulation) was not standardized in the final models.

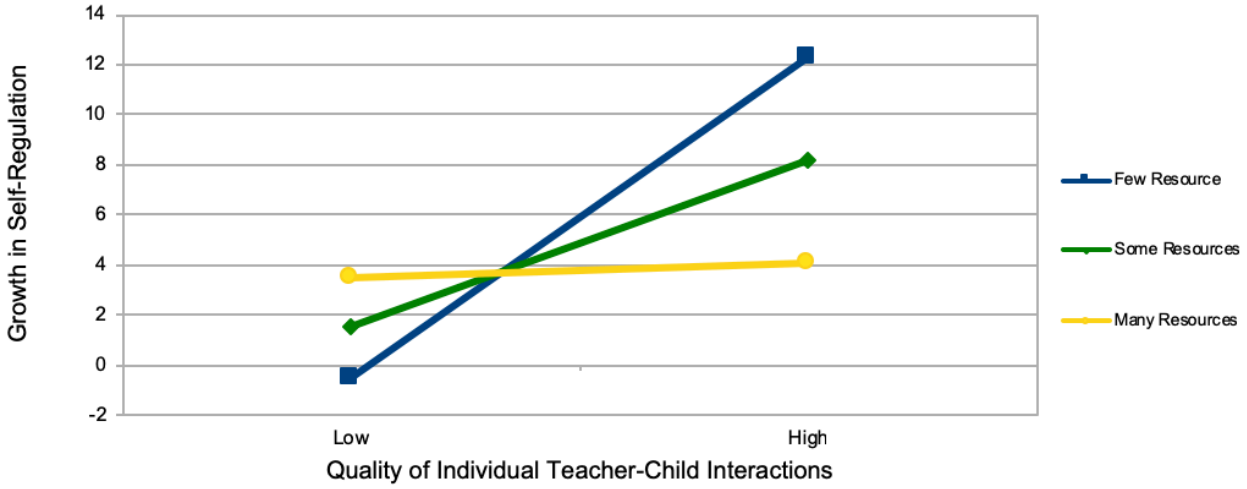

Figure 1. Combined effect of the quality of individual teacher-child interactions and neighborhoods resources for basic and daily needs on growth in children's self-regulation.