2019

The Home Math Environment and Math Achievement: A Meta-Analysis

Mia Cristina Daucourt

FLORIDA STATE UNIVERSITY

COLLEGE OF ARTS AND SCIENCES

THE HOME MATH ENVIRONMENT AND MATH ACHIEVEMENT: A META-ANALYSIS

By

MIA CRISTINA DAUCOURT

A Thesis submitted to the
Department of Psychology
in partial fulfillment of the
requirements for the degree of
Master of Science

2019

Mia Daucourt defended this thesis on July 10, 2019.

The members of the supervisory committee were:

Sara A. Hart
Professor Directing Thesis

Colleen Ganley
Committee Member

Alexandria Meyer
Committee Member

The Graduate School has verified and approved the above-named committee members, and certifies that the thesis has been approved in accordance with university requirements.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Mathematical thinking is in high demand in the global market, but compared to their international peers, U.S. school children fail to meet math performance benchmarks. This is especially problematic, given that early math skills predict later success in math and reading, beyond the effects of early reading skills and that math difficulties prior to formal schooling make it unlikely that *children who start off behind will catch up*. The home math environment (HME), which includes all math-related activities, attitudes, expectations, resources, and interactions between parents and children in the home, provides a potentially promising way to promote children's early math development. In order to understand the role played by the HME in children's math abilities, the a pre-registered meta-analysis was conducted to estimate the average weighted correlation coefficient, r between the HME and children's math achievement and the sample, assessment, and study features that contribute to study heterogeneity. A multilevel correlated effects model was run on 51 studies and a total of 456 effect sizes, which found a positive, significant average weighted correlation of $r = .14$, $p < .0001$. Although the association found was low in magnitude, our combined sensitivity analyses showed that the present findings were robust, and that the sample of studies has evidential value. Interestingly, moderator analyses revealed that all moderators tested contributed to study heterogeneity and when the HME component moderation analyses were run, no significant between-study heterogeneity remained.

Keywords: home math environment; math; meta-analysis

CHAPTER 1

INTRODUCTION

The Home Math Environment and Math Achievement: A Meta-Analysis

In an age when mathematical thinking has become integral to sustaining a competitive advantage in the global market, a critical national concern is the failure of U.S. school children to meet the same math performance benchmarks as their international peers (Provasnik Kastberg, Ferraro, Lemanski, Roey, & Jenkins, 2012). Given that math achievement deficits already exist at the onset of formal schooling, and children who start off behind in school are unlikely to catch up to their peers (Aunola, Leskinen, Lerkkanen, & Nurmi, 2004; Jordan, Kaplan, Locuniak, & Ramineni, 2007), school-based instructional efforts to improve math outcomes are likely not enough. Alongside research that has found a strong association, even beyond the effects of social class, between general home learning activities and student achievement (Bus, Van IJzendoorn, & Pellegrini, 1995; Kellaghan, Sloane, Alvarez, & Bloom, 1993), evidence also shows that children's early math knowledge develops when they have opportunities to engage with and talk about math in a playful, low-stakes manner (Cohrssen, Church, & Tayler, 2014). Thus, the home math environment (HME), which encompasses all math-related interactions among parents and children in the home, including informal board game playing, using words that compare magnitudes (i.e., "more" or "less"), and other low-stakes math-related exchanges, may provide a promising avenue for the development of children's early math skills before school entry.

Practical Significance of the HME

In terms of school success, early math skills have been shown to play a vital role in academic achievement. In fact, early math skills not only strongly predict later math achievement (Duncan et al., 2007) but have also been shown to predict later reading achievement more strongly than do early reading skills (Duncan & Magnuson, 2011). However, in comparison to research on the home literacy environment (HLE), research on the HME is much less abundant (e.g., Senechal & LeFevre, 2002). Given the confluence of evidence on the importance of the HLE in children's literacy skill development before school entry, it is reasonable to expect that variations in home math experiences may also be driving the large differences in children's early math skills prior to formal schooling (Aunola et al., 2004; Evans & Shaw, 2008; Huntsinger et al., 2016; Phillips, Norris, & Anderson, 2008; Senechal & LeFevre, 2002). Nevertheless, the role of the HME in children's math achievement remains unclear, with reported correlations between the HME and children's math achievement ranging from small to large and positive to negative.

HME Definitions

One of the problems that may be driving inconsistencies in the HME literature is the lack of agreement on how the HME should be defined. A common thread among the many conceptualizations of the HME is their emphasis on parent involvement with math (e.g., Niklas & Schneider, 2013), but there is no consensus on the specific components that ought to be used to capture this parent involvement. Overall, research examining the home influences on children's math achievement have ranged from single-factor definitions of the HME to a wide variety of multi-component definitions that include many different combinations of components, like explicit instructional activities involving math, everyday activities that incidentally involve math, math-related resources, like board games and number books, parent math talk, and parent affective factors and beliefs related to math, like math attitudes (e.g., math anxiety) and math achievement expectations (Benavides-Varela et al., 2016; Ciping, Silinskas, Wei, & Georgiou, 2015; Kleemans, Segers, & Verhoeven, 2015; del Rio et al., 2017; LeFevre et al., 2009; LeFevre, Polyzoi, Skwarchuk, Fast, & Sowinski, 2010; Ramani, Rowe, Eason, & Leech, 2015). Given that there is a wide array of home-based factors potentially encompassed by the HME, and the HME has been defined differently among different research areas and studies, it is not surprising that the findings on the association between the HME and child math achievement are also inconsistent (e.g., Blevins-Knabe, 2000; Ciping et al., 2015; Huntsinger, Jose, & Luo, 2016).

Across many empirical studies on the HME, the HME has been operationalized narrowly, encompassing only the (purposefully- and incidentally-) math-related activities that parents and children share in the home, but more recent empirical work has expanded the way the HME is defined to also incorporate parental cognitions about math (Skwarchuk et al., 2009; Skwarchuk, Sowinski, & LeFevre, 2014). Parental cognitions include parents' math-related attitudes (i.e., math anxiety, math importance), as well as parents' expectations for their children's math achievement and have been shown to directly influence parenting practices as well as children's math attitudes and outcomes (Cheung, Yang, Dulay, & McBride, 2017; del Rio, Susspereguy, Strasser, & Salinas, 2017; Kleemans, Peeters, Segers, & Verhoeven, 2012; Taylor, Clayton, & Rowley, 2004). Importantly, compared to other forms of parent involvement, parental aspirations and/or expectations have been shown by meta-analytic work to have the strongest association with children's academic achievement (Fan & Chen, 2001). Moreover, according to Eccles' Theory of Parent Socialization, parental attitudes work as an important social reinforcer of children's beliefs about their (math) abilities and math, in general, which in turn, impact children's academic (i.e., math) performance overall (Eccles, Jacob, & Harold, 1990). Thus, parent math-related attitudes and expectations are an important consideration when measuring the influence of the HME on children's math achievement.

One more line of inquiry that has emerged in the study of home and parental influences on children's math achievement, which has not traditionally been conceptualized as part of the

HME, is parent number talk, or "parent math talk". Parent math talk refers to parent utterances of number words (i.e., one, two) and words related to magnitude comparisons (i.e., more, less; Gunderson & Levine, 2011; Levine, Suriyakham, Rowe, Huttenlocher, & Gunderson, 2010). Although the amount and frequency of math talk parents engage in with their children has been shown to predict children's later math achievement (Levine et al., 2010), investigations of parent math talk in the home have yet to be combined with traditional HME studies that focus on math-related activities, attitudes, and expectations in the home. This is surprising, given that all these math-related influences are present in the home environment at once, and are likely exercising a significant influence on the math-related social interactions between parents and children. In light of this evidence, the current meta-analysis was conducted with a comprehensive definition of the HME that combined parent- child math activities as well as parent math attitudes, parent math expectations, and parent math talk. Importantly, this will allow us to more fully capture the HME and empirically test whether math talk is related to children's math achievement across studies.

Theoretical Basis of the Link Between the HME and Children's Math Achievement

A large body of research on the development of mathematical knowledge shows that numerical ability develops from infancy through adolescence (e.g., Geary, 1994). This math development spans from the innate numeracy skills we are born with and to formal school-based math learning. In between innate math skills and the math skills children learn in school, another mechanism for math skill development to consider is home-based math learning. In fact, the HME represents a setting in which many interacting influences converge to affect children's outcomes, including children's individual characteristics, environmental factors (Bronfenbrenner, 1979), social interactions (Vygotsky, 1978), and cultural influences (Bornstein

& Cheah, 2006).

In line with the view that environments help shape children's development, Bronfenbrenner's bioecological systems theory (1979) proposes that children's academic achievement and social development are joint functions of children's individual characteristics and their interactions with two levels of contextual influences—proximal and distal. Proximal contexts directly involve the child, like his/her home, family, and early childhood programs, and distal contexts are those that indirectly involve a child, like his/her parents' workplace. In this framework, the HME represents a proximal microsystem, in which parents directly influence their children's math skill development through the provision of math-related activities, attitudes, utterances, and resources within the home. In turn, children's math abilities also exercise a complimentary effect on their math-related home environmental influences. Thus, the HME represents a setting that captures the bioecological interaction between children's math skills and their math-related home environment, which is likely to directly influence children's math achievement.

One mechanism by which the HME may exercise a direct influence on children's math achievement is through parent-child social interactions within the home. According to Vygotsky's sociocultural learning theory (1978), children's cognitive abilities develop through social interactions with more experienced partners, which push children toward an upper boundary of ability that they could not reach on their own. Within the HME, these social interactions may include parent-child math-related activities and utterances as well as the socialization of math attitudes through talking about feelings toward math and expectations for children's achievement in math. As such, the math-related social exchanges that parents share with their children as part of the HME may lead to a level of math learning that surpasses what children could achieve without such math-related interactions. It stands to reason that the social exchanges within the HME may be partially driving the differences in math ability found in young children prior to formal schooling (Aunola et al., 2004). Thus, the present meta-analysis will directly investigate the influence of the different kinds of social interactions within the HME in order to examine the association of HME interactions with children's math achievement.

An important over-arching influence on the individual, social, and environmental factors within the HME is culture. In fact, within the HME, culture drives variations in physical and social settings, as well as in child-rearing customs and practices (i.e., frequency of interaction), and in the psychology of caregivers (i.e., level of strictness versus nurturing; Harkness & Super, 2002). According to Bornstein and Cheah's (2002) conceptualization of the home environment, the HME represents a developmental niche in which culture shapes the beliefs, attitudes, and values parents pass on to their children about math, which in turn, impact how children's math-related knowledge and abilities develop. Thus, the HME may represent a vital context in which many influences on children's math development converge to impact their math achievement. Given the environmental, social, and cultural influences captured by the HME, a statistical examination of the role of the HME in children's math achievement is an important empirical question.

Study Aims

In response to the lack of consensus on how the HME is measured and defined, and the inconsistent findings in the literature on the magnitude and direction of the relation between the HME and children's math achievement, the current, pre-registered meta-analysis was conducted to synthesize all available empirical evidence on the HME-math achievement relationship in order to accomplish two over-arching goals. We preregistered the design, research questions, moderators, and analysis plan of this systematic review with Prospero under the title "A meta-analysis on the relation between the home math environment and children's math achievement" and the registration number CRD42018099626. Our first study aim was to combine all previous studies conducted on the association between the HME and children's math achievement in order to utilize the statistical power of a combined sample of empirical studies to calculate the average weighted correlation between the two constructs. Our second study aim was to empirically test a range of potential moderator variables that may influence the strength of the relation between the HME and children's math achievement in order to determine the between-study sample, measure, and study characteristics that contribute to the differences found across studies.

Literature Review

We first consider the existing literature that links the HME to children's math achievement. Following that, we examine a range of potential moderator variables that may influence the relation between the HME and children's math achievement.

Inconsistent Findings on the Association Between HME and Children's Math Achievement

Overall, previous work examining the relation between the HME and children's math achievement has yielded inconsistent findings on the magnitude and direction of their association. A number of studies have found a positive and significant correlation between the HME and math achievement in children ranging from pre-school to elementary school, whether math was measured concurrently or longitudinally (Benavides-Varela et al., 2016; Dearing et al., 2012; del Rio et al., 2017; Hart et al., 2016; Huntsinger et al., 2016; Kleemans et al., 2012; Kleemans et al., 2013; LeFevre et al., 2009, 2010; Manolitsis, Georgiou, & Tziraki, 2013; Niklas, Cohrssen, & Tayler, 2016; Segers et al., 2015; Skwarchuk et al., 2014). For example, a handful of studies that included parents' home numeracy activities and parent numeracy expectations in their definition of the HME found that each component uniquely predicted children's math achievement across a spectrum of early numeracy skills, in cases where math was assessed on standardized or unstandardized tests (del Rio et al., 2017; LeFevre et al., 2010; Kleemans et al., 2012, 2013; Niklas et al., 2016; Segers et al., 2015). Thus, despite differences in the studies' sample age, math assessments used, or conceptualization of the HME, the association between the HME and children's math achievement was still found to be positive and significant.

Conversely, many studies have also found either a negative or non-significant association between the HME and children's math achievement (Ciping et al., 2015; Huntsinger et al., 2016; Missall, Hojnoski, Caskie, & Repasky, 2015). Negative associations with children's math achievement have been found across a variety of samples (i.e., different countries of origin), math assessments (i.e., timed versus untimed), and HME components (i.e., activities that target math on purpose versus incidentally; Ciping et al., 2015; Huntsinger et al., 2016). The same diversity of study characteristics holds true for null findings in the association between the HME and children's math achievement. Indeed, null findings have been found with math assessed using a variety of school readiness and numeracy skill measures, and with diverse sample characteristics (i.e., low SES, ethnic minority preschoolers or middle school children; Leyva et al., 2017; Missall et al., 2015). It may be the case that the negative and non-significant findings for the association between the HME and children's math achievement are an issue of directionality. Specifically, it may be the case that children struggling with math are more likely to need help with math, and thus share more math-related interactions with their parents at home (Ciping et al., 2015). Indeed, parents may intentionally focus on more basic skills when their children struggle to meet developmental math demands (Saxe, Guberman, & Gearhart, 1987). No matter what is driving these conflicting findings, there is more work to be done to understand why the association that exists between the HME and children's math achievement is so variable between studies. As such, the moderator analyses conducted in the current study will enable us to statistically test for and pinpoint the study characteristics that are driving between-study variability.

Potential Moderators of the HME-Children's Math Achievement Link

A number of factors may have contributed to the inconsistent findings in previous research on the HME and children's math achievement. These factors include sample characteristics—country of origin, age, grade, and special population characteristics, HME assessment methods—the HME component(s) measured, the category of the HME component measured (i.e., attitude, activity, or math talk), and how the HME score was calculated, math assessment methods—the math assessment used, the math domain measured, and whether or not the math measure was symbolic, timed, standardized, or a composite, or study characteristics—whether or not the study was longitudinal. Table 1 shows the specific coding scheme for all moderators included in the moderator analyses.

Sample Characteristics.

Age and grade. In an effort to capture the effects of the HME prior to formal schooling, the majority of HME research has been conducted on preschool and kindergarten samples. However, recent work on age-related differences in the HME has shown that more advanced HME activities are correlated with older children's, but not younger children's, math performance (Thompson, Napoli, & Purpura, 2017). This may indicate that the role of the HME in children's math achievement may differ depending on the age of the sample, with more advanced activities being more important for older children's math performance and more basic activities being more important for younger children's math performance. However, given that basic HME activities were not related to younger or older children's math achievement, when SES was controlled for, the expected pattern of results was not supported. Moreover, both younger and older samples have shown both positive and negative correlations between the HME and children's math achievement, further indicating that age-related differences in the association between the HME and children's math achievement do not follow a predictable pattern (Ciping et al., 2015; LeFevre et al., 2009; Pezdek et al., 2002). Given that there appears to be no clear direction for the potential moderating effect of age, moderator analyses will be conducted for both age and grade to figure out how age and grade may influence the correlation between the HME and children's math achievement.

Country of origin. Cross-country work has shown parent attitudes and practices directly influence children's math performance and that parent achievement attitudes and practices differ depending on the parents' birth country (LeFevre et al., 2010; Huntsinger et al., 2016; Missall et al., 2015). For example, previous work has found that, comparatively, Chinese-American and Taiwanese parents are more likely to engage with their children in formal math activities and do so when their children are younger than their European-American counterparts (Huntsinger, Jose, Liaw, & Ching, 1997). Based on these findings, and the influence of culture on individual, environmental, and social factors in the home environment (Bornstein & Cheah, 2002), we included the study sample's country of origin as a potential moderator of the relation between the HME and children's math achievement. Specifically, we expect Asian countries to have higher magnitude correlations between the HME and children's math achievement than European countries and the United States, based on Asian countries' cultural focus on parent involvement, especially in the form of academic expectations (Huntsinger, Jose, Larson, Krieg, & Shaligram, 2000).

Special sample characteristics. Given that math achievement differences, which coincide with differences in socioeconomic status, race/ethnicity, and gender are evident in young children, even before the onset of formal schooling (Aunola et al., 2004; Gibbs, 2010; National Mathematics Advisory Panel, 2008), it is reasonable to expect that qualitative differences in the home environment between children of different genders, socioeconomic, and/or ethnic backgrounds, or other defining characteristics may be contributing to the differences found in children's math development (i.e., SLI; Baker, 2015; Cheung et al., 2017; Kleemans et al., 2013). In fact, both gender- and SES-based math achievement gaps have been found, with high SES samples and boys tending to perform better (Wei, Lenz, & Blackorby, 2013), which may also be an indication that the math achievement of high SES children and boys is more positively impacted by or closely related to the HME. Thus, in an effort to capture the impact of sample characteristics on the association between the HME and children's math achievement, we conducted moderator analyses for the special sample characteristics that varied between studies on the HME-math achievement relation.

Assessment Characteristics.

HME assessment: component measured and score calculation method used. As previously discussed, one of the most common differences among studies examining the HME and children's math achievement is the way researchers operationalize the HME. The most common components used to capture the HME include parents' math-related cognitions (i.e., attitudes and/or expectations), math-related activities shared between parents and children, or some combination of the two. Surprisingly, even though it represents another home-based mechanism of math-related social learning, parent math talk tends to be excluded from traditional HME definitions and studied as a separate line of research. In an effort to bring together HME research on parent math talk and more traditional HME research on parent math attitudes and activities under one HME umbrella, the present meta-analysis combined studies from both previously separate research literatures in order to calculate an average effect size that is based on a more comprehensive snapshot of the HME.

We also accounted for more fine-grained differences across HME definitions. These differences included whether direct math activities, indirect math activities, a combination of direct and indirect math activities, spatial activities, math expectations, math attitudes and/or beliefs, math talk, or a combination of math-related activities and attitudes and/or beliefs or expectations, or some other combination of two HME components was measured. Based on previous HME work showing that, even when children's math achievement is measured with the same math assessment or within the same math domain, different magnitude correlations are still found between the HME and children's math achievement when different aspects of the HME are measured (Skwarchuk et al., 2014; Thompson et al., 2017), we expect the HME component measured to be a significant moderator on the association between the HME and children's math achievement. Specifically, we expect effect sizes that measure the HME as indirect math activities or math attitudes and/or beliefs to have lower magnitude correlations with children's math achievement than those that measure direct HME activities (Huntsinger et al., 2016; Jacobs & Harvey, 2005). In addition, based on findings from seminal studies in the HME literature as well as meta-analytic work, we expect correlations that include the HME category of parent math expectations to have the highest magnitude correlations with children's math achievement compared to all other HME components (Fan & Chen, 2001; Skwarchuk et al., 2014).

Another way in which HME operationalizations differ is in the method used to calculate

HME scores. Many studies utilize exploratory factor analyses to create one or more latent factors representing the HME, but other studies utilized sum or average scores from the HME assessment items, or simply analyzed single HME questions. Based on the reduction in measurement error provided by latent factors in comparison to measured variables (Gayán & Olson, 2003), we expect to find higher magnitude correlations when the HME is measured as a latent factor, rather than either of the other two methods mentioned. Thus, we will test whether the HME category measured, the HME component measured, or the method used to calculate the HME score significantly moderates the association between the HME and children's math achievement.

Math assessment: assessment, domain (un)timed, (un)standardized, composite or single measure. The math achievement literature consistently shows that math ability is made up of many component skills that are related yet distinct (e.g., Purpura & Ganley, 2017). Accordingly, the majority of the research on the HME and children's math achievement has investigated a wide array of math skills and spatial skills (e.g., Dearing et al., 2012; Thompson et al., 2017). However, most HME research is conducted in young children, leading the majority of studies on the HME-math achievement link to be focused on informal numeracy skills, which can be grouped under the overarching categories of numbering, relations, and arithmetic operations (NRC, 2009; Purpura & Lonigan, 2013). Numbering refers to understanding of the rules and processes associated with counting, including verbal counting, counting errors, one-to-one correspondence, cardinality, subitizing, and estimation (Purpura & Lonigan, 2013). Numerical relations refers to understanding of the ways two symbolic numbers may be associated, including quantity comparison, number comparison, number naming, ordinality, and number line sequencing (Purpura & Lonigan, 2013). Finally, arithmetic operations refers to knowledge of the ways sets and subsets of numbers can be created and decomposed, including addition, subtraction, and other forms of combining numbers (Purpura & Lonigan, 2013). Based on evidence that different kinds of home math experiences significantly predict some math skill domains but not others (e.g., Benavides-Varela et al., 2016; LeFevre et al., 2009), and the majority of studies showing that the direction, strength, and significance of the association between the same HME component and children's math achievement differs between math domains (i.e., Dearing et al., 2012; Kleemans et al., 2013; Huntsinger et al., 2016; Missall et al., 2013; Yildiz et al., 2018), we believed that the math domain measured may be driving the different correlations found between the HME and children's math achievement. For example, the negative association between the HME and children's spatial skills may be attributable to measuring the spatial skills domain, rather than any differences in the HME measurement (e.g., Susperreguy et al., 2018; Zippert & Rittle-Johnson, 2018). However, since some studies have also found evidence to the contrary, with correlations of similar magnitudes between the HME and children's math achievement across different math domains (Vukovic et al., 2013), we tested whether math domain was a significant moderator of the association between the HME and children's math achievement directly. In order to ensure that the potential differences found were attributable the math domain being measured, and not to the specific math assessment, we also accounted for the math assessment used as well as a number of specific assessment characteristics.

Moderator analyses for specific math assessment characteristics, including whether the math assessment was symbolic, non-symbolic, or both, standardized or unstandardized, timed, untimed, or both, or whether math achievement was assessed using a composite or single measure were also run. These characteristics were chosen based on the math assessment features that were found to vary most frequently between HME studies, and based on etiological differences found in genetically-sensitive literature (Hart, Petrill, Thompson, & Plomin, 2009; Petrill et al., 2012), which may indicate that different levels of environmental input are present in the HME based on different math assessment features. Thus, we also tested whether the math assessment used and its assessment features moderated the correlation between the HME and children's math achievement.

Study Characteristics.

Study characteristics: longitudinal. Whether or not a study captured longitudinal (at different time points) or concurrent (at the same time point) relations between the HME and children's math achievement may also moderate the correlation found. Specifically, it is possible that the HME has a stronger relation with math achievement depending on whether math was assessed concurrently or later on (i.e., a year or two later). For example, the benefits conferred by the HME for children's math achievement may weaken over time (Manolitsis et al., 2013). On the other hand, the effects of the HME may take time to be reflected in children's math performance, resulting in stronger effect sizes for longitudinal associations compared to studies that measure the HME and children's math achievement concurrently. Given this possibility, we will account for whether or not a study included an effect size that captured longitudinal or concurrent relations in our moderator analysis.

CHAPTER 2

METHODS

Literature Search

To begin, databases to be searched were chosen based on running a search (“home numeracy environment” AND parent* AND home) in EBSCO Discovery Science, a search tool that draws from all university-accessible databases, and annotating which databases came up in the results. Two expanders were added to this search, namely “also search within the full text of the articles” and “apply equivalent subjects.” The databases included: Education Source, Academic Search Complete, Education Full Text (H.W. Wilson), Social Sciences Citation Index, Academic OneFile, Education Resources Information Center (ERIC), Child Development & Adolescent Studies, MEDLINE (PubMed), MEDLINE (ProQuest), PsycARTICLES, PsycINFO (including PsycINFO Theses and Dissertations), and Social Sciences Full Text. All of these databases were searched one by one with the comprehensive search terms: ("home math environment" OR "math talk" OR home OR "home environment" OR "home learning" OR "home experience" OR "home numeracy" OR "informal learning environment" OR "home practices" OR "home activities") AND ("parent child interactions" OR "parent school relationship" OR "parent characteristics" OR "parent expectations" OR "parents as teachers" OR "parent student relationship" OR "parent child relations" OR "parent attitudes" OR "parent beliefs") AND ("number activities" OR "number skills" OR numeracy OR "early numeracy" OR math* OR "math skills" OR "math ability" OR "mathematical reasoning").

Next, additional databases were chosen by reviewing the databases listed on Florida State University’s library research guides for related topic areas in Psychology, Mathematics, Education, Early Childhood Education, and Family and Child Sciences. Based on these guides, three additional databases were included: Educators Reference Complete, MathSciNet, Web of Science. In order to be as comprehensive as possible and capture the grey literature that may not be found in topic-specific databases, Google Scholar was also searched.

Once the database searches were all conducted, the results were saved in RefWorks to be reviewed for inclusion in the present meta-analysis. First articles were excluded based on duplicates (which RefWorks automatically detects). Second, all non-duplicates were reviewed to determine whether studies met inclusionary criteria based on titles and abstracts, and in cases where it was necessary, a review of methods, tables, and/or full manuscripts was also conducted to determine whether inclusionary criteria were met.

Once the final sample of articles was determined, reference lists were then reviewed in order to determine whether there were any articles cited by an included article that did not show up in the manual searches. The references were also reviewed to identify prominent authors in the area that had multiple publications or a seminal publication in the HME research area (i.e., a paper that was highly-cited, had methods that were highly replicated, or developed a frequently-used HME measure). Then, the Google Scholar profiles of prominent authors in the area were reviewed to make sure all of their relevant work that was not included in the study sample from manual searching was captured. Finally, although Dissertations and Theses were included in the database search results (especially Google Scholar), and also directly captured by searching PsycINFO Theses and Dissertations, a p-curve analysis was conducted to statistically test for the presence of publication bias.

Inclusionary and Exclusionary Criteria

To be included in the present meta-analysis, a primary study had to meet the following criteria:

1. A study must have an operationally-defined HME measure. The HME must measure practices, attitudes, expectations, and/or beliefs that are math-specific separately from other achievement domains (e.g., literacy, science) or the study will be excluded. For example, a study that only used a general home learning environment measure (e.g., Casey et al., 2014; Crosone et al., 2010; Hindman et al., 2010; Foster et al., 2016; Galindo & Sheldon, 2012), without separating math-specific aspects, would not be included. Studies that measure home math talk (e.g., Ramani et al., 2015) and parent attitudes, beliefs, and/or expectations toward math (e.g., del Rio et al., 2017; Segers et al., 2015) will also be included in our conceptualization of the HME, but only if they are math-specific. Informal home-based play activities that involve math, like video games, board games (e.g., Benavides-Varela et al., 2016; Huntsinger et al., 2016), and grocery games (e.g., Pezdek et al., 2002) will also be included as part of the HME under the category of indirect math activities.
2. A study must include at least one math-specific achievement measure that does not include other achievement domains (e.g., language skills, Keith & Lichtman, 1994) in order to enable us to isolate the effect of the HME on math achievement only. The math achievement measure can involve any assessment method, including parent-report of children's math skills (Hart et al., 2016), and standardized (e.g., Blevins-Knabe & Musun-Miller, 1996; Cheung et al., 2017) and unstandardized (e.g., LeFevre et al., 2009; Skwarchuk et al., 2014) math tests. Studies that examine the HME but have no math achievement measure will be excluded (e.g., Anderson, 1997; Missall et al., 2017).

3. If a study reported more than one math achievement outcome, the same math achievement outcome at multiple time points, and/or has more than one component of the HME included, the multiple combinations of HME measure and math achievement measure will each be included as separate effect sizes. Then, the effect sizes will be pooled, while accounting for dependent effect sizes by using multilevel correlated effects models to control for study effects using R's metafor package (Viechtbauer, 2010).
4. If a study did not report the zero-order correlation between the HME and a math achievement outcome, did not report sufficient statistics to allow us to derive a zero-order correlation between the two, then the primary corresponding author of the study will be contacted via E-mail in an attempt to procure the missing information. If the author does not respond within two weeks or chooses not to provide the information we need, the study will be excluded. A consequence of this criterion is that only quantitative examining the HME-math achievement relation will be included from the present analysis and all qualitative studies will be excluded.

Coding Procedures and Reliability

For the present meta-analysis, we implemented a systematic process for identifying and coding the study results (i.e., Pearson correlation coefficient and corresponding sample size) and study descriptors (i.e., moderators) from the primary studies (outlined in Table 2). The coding was done in three phases by two authors of this article. The first "trial coding" phase began with discussion between the authors of this paper, as well as their graduate student advisors, which have expertise in the environmental and affective factors influencing math achievement and experience in meta-analysis, to create a tentative coding plan. Additionally, the instructor of a graduate-level meta-analysis course, who is an expert in meta-analytic methodology, was consulted on the creation of the tentative coding plan (as well as the methodological choices outlined in the following sections) as part of a final assignment for class credit. In addition, coding features were amended and finalized based on the authors' and their advisors' reading of relevant articles in the research area.

The initial coding phase also involved the coding of five primary studies by both coders, followed by a comparison of their coding consistency and discussion of any issues that needed clarification or verification. The inter-coder reliability had to be high (Cohen's kappa equal to or greater than .75, Fleiss, 1981; Cicchetti & Sparrow, 1981) before moving onto the second phase. The second coding phase involved eliminating articles that were identified as irrelevant or did not meet the inclusionary criteria until the total sample of studies was finalized. As stated previously all articles were reviewed and excluded based on duplicates, followed by a review of titles and abstracts, and a then a final review of methods, table, and possibly full articles in order to determine if each study met inclusionary criteria.

Coding Procedures. The coding scheme included: Author(s), year of publication, average grade of the sample, average age of the sample, country of origin of the sample, special population sample characteristics, HME assessment component, and calculation used, math assessment used, math domain assessed, whether or not the math assessment used was a standardized, timed, symbolic, or a composite, and whether or not the study was longitudinal (see Table 2). A column was also created to code for correlations that came from the same article and study sample (i.e., study ID). These final data were then imported into R, and all Pearson's correlation coefficients, *r*, and the corresponding sample sizes, *n*, were used to calculate the corresponding Fisher's Z and variance for each effect size using the escalc() function from R's metafor package. Then, Fisher's z-transformed effect size was utilized for all subsequent analyses in R.

Age and grade. Because children's grades are not always translatable across countries (e.g., kindergarten starts a year later in Norway compared to the U.S.) and not all manuscripts report the age of their study sample, both grade and age were utilized to ensure more complete data and to more precisely capture the potential moderating effects of a child's point in development on the relation between the HME and children's math achievement. In cases where age or grade was not reported, just the sample feature that was reported in the manuscript was coded. In the cases where age or grade was reported as a range, the average was calculated and subsequently used. Specifically, grade was coded as 1 = preschool and/or kindergarten (PK/KG), 2 = a combination of PK/KG and elementary school, 3 = elementary school, 4 = a combination of elementary and middle school, and 5 = middle school. The grade categories were named according to the United States school system. Thus, elementary school included grades 1 through 5, and middle school included grades 6-8.

Special sample characteristics. Specifically, we coded the sample as 1 = average or typically-developing, 2 = low SES (30% or more of sample low SES), 3 = high minority (30% or more of sample made up of Non-white, minority participants), 4 = all one ethnicity (75% or more of sample from one country or ethnic origin), 5 = all girls (100% of sample), 6 = all boys, (100% of sample), 7 = sample with Specific Language Impairment (30% or more of sample made up of SLI participants), 8 = had some other special selection criterion, or 9 = high SES (30% or more of sample high SES).

HME component. The HME component was coded as 2 = direct activities, 3 = indirect activities, 4 = attitudes and/or beliefs, 5 = math expectations , 6 = spatial activities, 7 = math talk, 9 = direct and indirect activities, and 11 = combination of activities and attitudes and/or beliefs or expectations. Direct activities were any math-related parent-child interaction that directly targeted math skills, like using flash cards, helping with math homework, or counting with children. Indirect activities captured any math-related parent-child interaction that incidentally targeted math skills, like cooking or playing board games. Math attitudes and/or beliefs included any parent affective factors toward math, like math anxiety and math importance. Math expectations represented parent expectations for their children's math achievement. Spatial activities were non-math activities that directly targeted spatial skills, like doing puzzles. Math talk included any math-related utterance made by a parent when interacting with their children, whether it included a comparison of quantity (i.e., more, less) or counting the squares aloud as they played board games.

Although playing board games was included in the indirect math activities category, the difference for including board game play in math talk was the inclusion of a score for the number of specific utterances parents made during board game play, not just an indication of whether or not parents played board games with their children. Thus, if an effect size included playing board games only, without tallying the number of specific parent utterances related to math during board game play, that correlation would be coded as an indirect activity, rather than math talk. On the other hand, if there was no measurement of parent speech during board game play, board game play would be coded as an indirect activity. A combination of direct and indirect activities represented parent-child math-related interactions that combined activities that both directly and incidentally targeted math skills. Finally, a combination of activities and attitudes and/or beliefs or expectations was used when more than one distinct component of the HME was included in a single HME score, like the inclusion of direct math activities and parent math anxiety in a single sum or latent factor score.

HME calculation. The HME calculation was coded as 1 = latent factor, 2 = sum score, or 3 = single HME item. A correlation as coded as a single HME item when the effect size included only a single question from an HME measure.

Math assessment. Based on the math assessments that were found most frequently in the HME literature, the math assessments were coded as 1 = researcher-created, 2 = KeyMath-3 Diagnostic Assessment (KeyMath), 4 = Test of Early Mathematics Ability 2 and 3 (TEMA), 5 = Woodcock Johnson-III/IV Tests of Achievement (WJ), 6 = Woodcock-Muñoz Batería III (WM), 7 = Utrecht Early Numeracy Test-Revised (UENT-R), 8 = Preschool Early Numeracy Skills test (PENS), 9 = California Achievement Test Mathematics subtest (CATM), 11 = Child Math Assessment (CMA), 12 = Other (including Individualized Growth and Developmental Indicators of Early Numeracy [IGDIs-EN] , Bracken Basic Concepts Scale- 3rd Edition: Receptive [BBCS:3-R] and School Readiness Composite [BBCS:3-SRC], Performance Indicators in Primary School [PIPS], Early Childhood Longitudinal Study-Kindergarten Cohort Math [ECLS-K Math], The Researcher-Based Early Mathematics Assessment Short-Form [REMA-S], Diagnostic Test for Basic Mathematical Concepts [DTBMC], Stanford Diagnostic Mathematics Test, Fourth Edition [SDMT4] Computation subtest), 13 = Multiple math assessments, 14 = Parent-report, or 18 = Test for Diagnostic Assessment of Mathematical Disabilities (TEDI-MATH). Multiple math assessments was used for an effect size only when the single effect size included more than one distinct math assessment but not if the effect sizes included multiple subtests from the same math assessment.

Math domain. Specifically, we coded math domain as 2 = arithmetic operations, 3 = numerical relations, 4 = numbering, 10 = multiple math domains, or 11 = spatial skills to examine whether the math domain measured contributed to study heterogeneity. These different math domains were previously described in the introduction.

Other math assessment characteristics. We coded for several nuances in how math was assessed in the HME literature, including whether the math assessment used was symbolic, with 1 = symbolic, 2 = non-symbolic, 3 = both symbolic and non-symbolic; timed, with 1 = timed, 2 = untimed, 3 = combination of timed and untimed; a composite, with 1 = composite of many math measures, 2 = a single measure; or standardized, with 1 = standardized, 2 = unstandardized, 3 = combination of standardized and unstandardized. Notably, only math assessments that included multiple different math assessments at once, rather than multiple subtests from the same math assessment were coded as a composite for the composite moderator and as multiple for the math assessment moderator. When math was assessed using multiple subtests from the same math assessment, the specific math assessment used was coded for the math assessment moderator, and single math measure was coded for the composite moderator.

Longitudinal study. Whether the study captured longitudinal (at different time points) or concurrent (at the same time point) relations between the HME and children's math achievement was coded as 1 = longitudinal, 2 = concurrent.

Coding reliability. The quality of the coding was evaluated by inter-rater reliability testing using Cohen's kappa (Cohen, 1960) on a random selection of 20% of the articles from the final sample (n = 10). Once a Cohen's kappa indicating a high inter-rater reliability (greater than.75), is achieved the coding will be deemed satisfactory and ready for analysis.

Effect Size Computation and Combining Effect Sizes

In the present meta-analysis, we examined the average association between the home math environment and children's math achievement, using the zero-order correlation coefficient, or *r* effect size. This *r* effect size was chosen because the empirical work examining the relation between the HME and children's math achievement uses primarily correlational designs that report Pearson correlations. A handful of studies in the final sample used experimental designs that employed HME interventions (e.g., Cain-Caston, 2013) and compared the math performance of those who participated in the intervention to control children's. In these cases, we reported only concurrent effect sizes that were calculated before the intervention was implemented. Once the *r* correlation coefficient(s) between the HME and children's math achievement were coded for each study, the effect sizes were converted using Fisher's *Z* transformation (Lipsey & Wilson, 2001). Specifically, each *r* was weighted by using the weighted w_i, which was based on the sample size associated with the *r* relative to the total accumulated sample size (see formula below):

$$Z_i = 0.5Ln\left(\frac{1+r_i}{1-r_i}\right),\ SE_i = \frac{1}{w_i}, (where\ w_i = n-3)$$

Then, the average Fisher's $\bar{Z}$ was obtained as the weighted average Z_i, using:

$$\bar{Z} = \frac{\sum(w_i \times Z_i)}{\sum w_i}, (where\ w_i = n-3)$$

Handling variability in effect sizes across studies. Once $\bar{Z}$ was obtained, the average weighted correlation between the HME and children's math achievement was calculated using a random effects meta-analysis. A random effects model assumes that there is a distribution of potential effect sizes that come from different populations, while a fixed effects model attributes study heterogeneity solely to sampling error, assuming that one universal effect size exists that comes from a single population (Borenstein, Hedges, Higgins, & Rothstein, 2009; Card, 2011). For the present analysis examining the relation between the HME and children's math achievement, a random effects model was chosen because the inconsistent methodology and definitional criteria used in HME research, and not just sampling error, most likely contribute to the high variability in effect sizes found. In addition, many of the studies included in the present analysis were conducted across a variety of different settings, including different countries. As such, it would be reasonable to assume that differences between studies represent true differences among different populations, rather than assuming that all study samples belong to a universal sample.

In order to support the choice of a random effects model, we evaluated the existence of heterogeneity statistically by conducting a *Q* test and calculating an unweighted sample-based estimate of I^2. The *Q* statistic tests for the presence of significant study heterogeneity based on a χ^2 distribution with *k*-1 degrees of freedom (Card, 2011; *k* = number of effect sizes used for the test). In order to show support for the use of a random effects model, the *Q* value must be greater than the critical value of the given degrees of freedom (i.e., have a significant *p*-value), indicating that the effect sizes are heterogenous and not attributable to sampling error. This also supports the use of follow-up moderator analyses should to account for the study features that contribute to effect size heterogeneity. On the other hand, if the critical value of the given degrees of freedom is smaller than the *Q* statistic, heterogeneity in effect sizes is considered to be non-significant, and the modeling may be adjusted to fixed effects. The *Q* statistic will then be converted into I^2 in order to calculate the proportion of variance in effect sizes, from 0 to 100%, that is attributable to heterogeneity:

$$I^2 = 100\% \times \frac{(Q - df)}{Q}$$

The use of both of these analyses will help empirically support our choice to use a random effects model and to conduct moderator analyses to determine the sources of study heterogeneity outside of sampling error.

Accounting for dependent effect sizes. In order to statistically account for the reporting of more than one effect size per study sample, a multilevel correlated effects model, controlling for study, was conducted with R's metafor package (Viechtbauer, 2010). The multilevel correlated effects model accounts for dependent effect sizes by modeling the Level-1 (effect size) and Level-2 (sample) correlations using maximum likelihood estimation (Fisher & Tipton, 2015). The methodological decision to use multilevel correlated effects modeling was based on the fact that our final data sample included 684 effect sizes that were drawn from only 52 studies, making it likely that accounting for study-level influences was necessary. Rather than requiring the calculation of an average effect size for each study or the extraction of covariance structures, the multilevel correlated effects analysis clustered dependent effect sizes by a given control variable (study ID) and weighted them based on correlated effects (rho or ρ), resulting in unbiased standard error estimates. The default value for correlated effects models of $\rho = 0.8$ was used for the current analyses. Then, in order to evaluate the robustness of our estimates, sensitivity analyses were conducted to test whether varying values of within-study correlations, or ρ (rho = 0.0, 0.2, 0.4, 0.6, & 0.8) impacted the values estimated for effect sizes, standard errors, and τ^2 (Hedges et al., 2010).

Evaluation of Publication Bias

Publication bias, which has become an increasing problem in the psychological sciences refers to the increased likelihood of studies with significant findings to be published and of studies with non-significant findings to be filed away in a drawer (i.e., the "file-drawer problem"; Rosenthal, 1979). Problematically, publication bias may lead to the estimation of a meta-analytic effect size that is smaller (or larger) than the true population effect size. Thus, in order to evaluate whether average weighted correlation between the HME and children's math achievement calculated for the current meta-analysis showed evidence of publication bias, R's metafor package (Viechtbauer, 2010) was used to conduct multiple tests using both visual and statistical techniques.

As an additional step, a *p*-curve analysis was also conducted to determine if there was evidence of *p*-hacking. P-hacking refers to the phenomenon where researchers collect or select data or modify statistical analyses until non-significant results become significant (Head, Holman, Lanfear, Kahn, & Jennions, 2015). Importantly, evidence of p-hacking typically indicates that a file-drawer problem exists because authors are likely to resort to p-hacking in order to obtain significant results so they can get their results published, while other researchers that do not p-hack and have non-significant findings are likely to be rejected for publication and filed away. We utilized p-curve analyses to determine the potential existence of publication bias by examining the distribution of significant *p*-values that corresponded to our observed effect sizes. P-curve analyses start with the calculation of pp-values, which represent the probability of obtaining each *p*-value if the null hypothesis (i.e., no significant effect) were true. These probabilities are then summed to derive a χ^2 value for testing the significance of the p-curve skew. A flat p-curve indicates that the probability of observing all p-values is uniform, and a right-skewed p-curve indicates that the effect is likely to be real, and the probability of lower p-values is greater than high p-values. Both of these scenarios likely point to a low chance of publication bias. However, a left-skewed p-curve shows evidence of p-hacking and indicates that the probability of high p-values is greater than the probability of low p-values. P-curve calculations were conducted within the p-curve application available at: http://www.p-curve.com/app4/, which provided both binomial and continuous tests for publication bias and p-hacking.

Funnel plot. Visually, publication bias was assessed using a funnel plot, a kind of scatter plot that visually depicts effect sizes relative to their standard errors (Card, 2011). A symmetrical distribution of observed effect sizes around the vertical line would indicate no publication bias, while an asymmetrical distribution would suggest potential publication bias. Symmetry was parametrically determined using the Egger test (Egger, Smith, Schneider, & Minder, 1997), which provides a *z*-estimate with an associated *p*-value that indicate whether or not asymmetry is significant.

Sensitivity Analyses

Fail-safe N. The Fail-safe N calculates the number of studies with null results (i.e., a statistically non-significant Pearson correlation coefficient) that would have to exist and be left out for the results of the average weighted effect size to be null. If evidence of significant publication bias is found, the metafor package will be used to make this calculation and determine the degree of publication bias.

Trim and fill. The 'trim and fill' method is used to correct funnel plot asymmetry by estimating what the results would be without publication bias and providing an estimate of the number of missing studies. The method estimates the true center of a funnel plot by removing the smaller studies driving funnel plot asymmetry and replacing the omitted studies and their missing 'counterparts' around the true funnel plot center (Duval & Tweedie, 2000a, 2000b). This analysis will also be conducted using metafor.

Robust Variance Estimation. In order to assess the robustness of our multilevel correlated effects estimates, R's robumeta package (Fisher, 2017) will be utilized to conduct a robust variance estimation analysis and follow-up sensitivity analysis across varying values of possible within-study correlations, or ρ (0.0, 0.2, 0.4, 0.6, & 0.8).

Excluding a Potentially Influential Study. In order to statistically determine if Cheung, 2017, a study with over 227 effect sizes had an inordinate influence on our meta-analytic results, we conducted a meta-analysis that excluded the study from the average weighted correlation calculation and compared the results to our average weighted correlation coefficient that included the study. If the effect sizes are found to be the same or similar, the study will be kept in our final study sample. If the effect size estimates are significantly different, the study will be excluded.

Data Analytic Plan

A meta-analysis was conducted to calculate the average correlation between the home math environment (which included home-based activities and talk related to math, parent attitudes and beliefs toward math, and parent math expectations) and children's math achievement (k = 52 studies, n = 684 effect sizes). Given that study samples spanned a wide range of ages, grades, and countries of origin, and a variety of home math environment and math achievement measures were utilized, a random effects model, which assumes that there is a distribution of true effect sizes instead of a single true effect size, was used to estimate the weighted average effect size, r (Hedges, 1983). For the purpose of comparison, a random effects model with all effect sizes was first conducted without controlling for sample dependence, but then a follow-up analysis was conducted in order to model clustering induced by effects derived from the same sample (Konstantopoulos, 2011; Nakagawa & Santos, 2012) utilizing multilevel correlated effects modeling (Viechtbauer, 2010).

Following the main analyses to calculate the average weighted *r* effect size between the HME and children's math achievement, a series of multilevel correlated effects moderator analyses, which controlled for study, were conducted to test sample, assessment, and study characteristics that may be driving study heterogeneity (Borenstein et al., 2009). Significant moderators were tested, one-by-one, with the metafor package based on the Q, I^2, and σ^2 statistics (Q_M; Borenstein et al., 2009; Higgins et al., 2003). The Q statistic and corresponding *p*-value indicated whether or not a significant portion of study heterogeneity was attributable to the given moderator (Borenstein et al., 2009), while the I^2 statistic captured the proportion of variance (0-100%) that was due to heterogeneity (Higgins et al., 2003), and the σ^2 statistic represented the true between-study variance from the observed studies (Konstantonopoulos, 2011).

CHAPTER 3

RESULTS AND DISCUSSION

Results

Included Studies

Our article searches yielded 1725 articles for review and coding. Only two articles from the final article sample were not captured by database searches and were procured from manual searching (Cai, 2003; Silinskas et al., 2010). During the first review of articles, 1190 of the articles were rejected based on titles or being duplicates (i.e., duplicates of both included and excluded articles), 431 more articles were excluded based on reviewing abstracts, and 52 were rejected based on reviewing methods or full manuscripts, resulting in a final sample of 52 articles. The article selection process is depicted in Figure 1. Given that 33% of the effect sizes from the final sample of 52 studies came from just one study (Cheung, 2013), that study was excluded from the main results and reported in Appendix A, resulting in 51 articles, reporting 456 effect sizes used in the main analyses. After rejecting articles based on titles that indicated that the association of interest was not included in the study (i.e., the article measured the home learning environment but did not include an achievement measure, the article was a review or a qualitative study, or the article measured children's reading instead of math achievement), the most common reason for article exclusion was the use of a home environment and/or achievement measure that was not math-specific. Once the final article sample was collected, the articles were divided between the two coders for extraction of data needed for effect size and moderator analyses. Upon completion of coding, a Cohen's kappa of .98 was achieved, indicating a high inter-rater reliability and that the data were ready for analysis.

Overall Average Weighted Correlation Between the Home Math Environment and Children's Math Achievement

Results of the random-effects analysis, which did not account for study dependence, yielded an average weighted correlation of .08 [.07-.10], *SE*=.01, $p < .0001$. However, given that only 51 studies resulted in a sample of 456 effect sizes, a multilevel correlated effects analysis was conducted to account for the large number of effect sizes drawn from the same study sample. The results from the multilevel correlated effects analysis yielded a higher average weighted correlation of .14 [.08, .19], $SE = .03$, $p < .0001$. Given the large difference in these estimates, with the results from the multilevel correlated effects analyses yielding an effect size that was nearly twice as large as the model that did not control for study, it appears that not accounting for study dependence drastically impacted our results, which provides support for the methodological decision to also account for study dependence when conducting follow-up moderator analyses.

Looking next at the results of the tests for study heterogeneity, significant heterogeneity was found, with Q [455] = 4279.97, $p < .0001$. The total heterogeneity of the r correlation coefficient was estimated to be high, $I^2 = 89.37\%$. Variance between studies was also found to be significant based on a 95% confidence interval, $\sigma^2 = 0.04$ [0.02-0.06]. Thus, multiple moderator analyses were conducted, one moderator at a time, in order to determine the sample, assessment, and study characteristics that may have significantly contributed to study heterogeneity.

Moderator Analyses for Sample Characteristics

All moderators were entered as categorical, except for age, which was entered as continuous. Age was a significant source of heterogeneity for the correlation between the HME and children's math achievement ($F[1, 428] = 7.95$, $p = .0050$, $\sigma^2 = .04$ [.02, .06], $k = 430$) with a

-.01 unit decrease in the correlation between the HME and children's math achievement for every 1-year increase in age. The test for residual heterogeneity was also significant (Q_E[428] = 1862.00, $p < .0001$), indicating that, even after accounting for age, variability in the observed effect sizes was significantly larger than would be induced by sampling error. In fact, I^2 = 77.01% of the variability in effect sizes was left unexplained after accounting for age.

All moderation analysis results, including the effect sizes, sample sizes, and 95% confidence intervals for each sample characteristic moderator and subgroup are presented in Figure 2. The results of the overall omnibus test with grade as the moderator showed that grade was a significant source of heterogeneity ($F(1, 410) = 8.06$, $p < .0001$, $\sigma^2 = .04$ [.02, .07], k = 415). However, only study samples that included only PK/KG children ($r = .15$ [.08, .21], $p < .0001$, $n = 298$) or only elementary school children ($r = .13$ [.06, .19], $p = .0002$, $n = 75$) had a significant influence on the correlation between the HME and children's math achievement, while study samples that included a combination of PK/KG and elementary school children ($r = .20$ [-.002, .40], $p = .0534$, $n = 30$), a combination of elementary and middle school children ($r = .05$ [-.21, .30], $p = .7236$, $n = 8$), and only middle school children ($r = .17$ [-.23, .57], $p = .4070$, $n = 4$) did not. When comparing differences between grades, the pairwise *t*-tests with PK/KG-only samples as the reference group showed that study samples comprised of only elementary school children demonstrated significantly lower correlations between the HME and children's math achievement than samples made up of only PK/KG children ($b = -0.02$ [-0.02,-0.01], $t(1) = -4.25$, $p < .0001$). The test for residual heterogeneity was also significant (Q_E[410] = 4097.53, $p < .0001$), and even after accounting for the grade of the study sample, I^2 = 89.99% of the variability in effect sizes remained to be explained.

The results of the overall omnibus test with country as the moderator showed that the study sample's country of origin was a significant source of heterogeneity ($F\,(10, 446) = 6.08$, $p < .0001$, $\sigma^2 = .02$ [.01, .04], $k = 456$) in the correlation between the HME and children's math achievement. Samples from the United States ($r = .08$ [.02, .14], $p = .0068$, $n = 250$), Canada ($r = .22$ [.09, .34], $p = .0006$, $n = 37$), the Netherlands ($r = .49$ [.33, .66], $p < .0001$, $n = 10$), Greece ($r = .23$ [.09, .38], $p = .0015$, $n = 11$), and Australia ($r = .26$ [.04, .49], $p = .0234$, $n = 6$) had average weighted correlations that were positive and significantly different from zero, but samples from Germany ($r = .09$ [-.22, .39], $p = .5799$, $n = 4$), Italy ($r = .05$ [-.25, .35], $p = .7480$, $n = 21$), China ($r = .10$ [-.03, .22], $p = .1263$, $n = 13$), Chile ($r = .05$ [-.16, .27], $p = .6202$, $n = 36$), and the other countries category ($r = .03$ [-.13, .18], $p = .7475$, $n = 68$) did not. When comparing differences between countries of origin, omnibus test results showed that study samples from the Netherlands ($b = 0.41$ [0.23, 0.58], $t(9) = 4.54$, $p < .0001$) demonstrated significantly higher correlations between the HME and children's math achievement than United States samples, but all other countries of origin did not significantly differ from United States samples. When samples from China were set as the reference group, contrary to our hypothesis, Chinese and U.S. samples did not significantly differ ($b = 0.01$ [-0.11, 0.13], $t(9) = 0.20$, $p = .8413$), and only Netherlands samples showed significantly different correlations, which were higher, rather than lower, than the correlation for Chinese samples ($b = 0.40$ [0.19, 0.60], $t(9) = 3.76$, $p = .0002$). The test for residual heterogeneity was also significant ($Q_E[446] = 4008.88$, $p < .0001$), and even after accounting for the country of origin of the study sample, $I^2 = 88.88\%$ of the variability in effect sizes was left unexplained.

Finally, the results of the overall omnibus test with special sample characteristics as the moderator showed that they were a significant source of heterogeneity ($F[9, 447] = 3.51$, p =

.0003, σ^2 = .03 [.02, .06], k = 456). Study samples that were average/typically-developing (r = .16 [.09, .24], p < .0001, n = 135), all one ethnicity (r = .12 [.03, .19], p = .0045, n = 148), all girls (r = .16 [.05, .26], p = .0041, n = 14), all boys (r = .15 [.04, .20], p = .0035, n = 13), or SLI (r = .54 [.17, .91], p = .0048, n = 4) had average weighted correlations that were positive and significantly different from zero, but samples that were low SES (r = .10 [-.01, .22], p = .0710, n = 96), high minority (r = .05 [-.32, .42], p = .7875, n = 18), high SES (r = .13 [-.25, .51], p = .4982, n = 7), or from the other category (r = .08 [-.04, .20], p = .1934, n = 21) did not. When comparing across special sample characteristics, omnibus test showed that in comparison to average/typically-developing study samples no significant differences in the average weighted correlation between the HME and children's math achievement were found due to special sample characteristics. Contrary to our hypotheses, low versus high SES samples (b = 0.03 [-0.37, 0.43], t(8) = 0.14, p = .8901), and samples that were made up of all boys versus all girls (b = -0.00 [-0.10, 0.09], t(8) = -0.04, p = .9678) did not show significantly different correlations between the HME and children's math achievement. The test for residual heterogeneity was significant (Q_E[447] = 4143.59, p < .0001), and even after accounting for special sample characteristics, I^2 = 89.21% of the variability in effect sizes was left unexplained.

Moderator Analyses for HME Assessment Characteristics

All moderation analysis results, including the effect sizes, sample sizes, and 95% confidence intervals for each HME assessment moderator and subgroup are presented in Figure 3. The results of the overall omnibus test with the HME component measured as a moderator showed that the specific HME component measured was a significant source of heterogeneity [F (8, 448) = 6.29, p < .0001, σ^2 = .01 (.01, .03), k = 456] in the average weighted correlation between the HME and children's math achievement. HME measures that assessed direct HME activities (r = .13 [.07, .19], p < .0001, n = 94), indirect HME activities (r = .08 [.02, .13], p = .0051, n = 155), parent math attitudes and/or beliefs only (r = .07 [.002, .15], p = .0446, n = 91), math expectations (r = .24 [.14, .34], p < .0001, n = 22), a combination of direct and indirect HME activities (r = .18 [.10, .26], p < .0001, n = 45), and a combination of activities and attitudes and/or beliefs or expectations (r = .20 [.05, .34], p = .0075, n = 15) had average weighted correlations that were positive and significantly different from zero, but HME measures that assessed spatial activities (r = .09 [-.08, .26], p = .2829, n = 8) and parent math talk (r = .09 [-.04, .23], p = .1812, n = 26) did not. When comparing between HME components, pairwise t-test results showed that, contrary to our hypothesis, when direct math activities was set as the reference group, no significant differences in the average weighted correlation between the HME and children's math achievement were found for direct versus indirect HME activities (b = -0.01 [-0.01, 0.04], t(7) = 1.06, p = .2890). Our pairwise t-test results with parent math expectations as the reference group partially supported our hypothesis, showing that only HME measures of parent math talk had a significantly lower magnitude average weighted correlations with children's math achievement than parent math expectations (b = 0.08 [0.04, 0.13], t(7) = 3.73, p = .0002). The test for residual heterogeneity was not significant (Q_E[448] = 409.98, p = .9007), indicating that when the specific HME component measured was accounted for, there was no significant variability in effect sizes left unexplained.

For the HME calculation moderator, the results of the overall omnibus test showed that the specific method of calculation used to measure the HME was a significant source of heterogeneity [F (3, 453) = 11.39, p < .0001, σ^2 = .04 (.02, .06), k = 456] in the average weighted correlation between the HME and children's math achievement. HME calculations that utilized latent factor scores (r = .14 [.08, .20], p < .0001, n = 136), sum scores (r = .16 [.10, .22], p <

.0001, $n = 215$), or single items ($r = .08$ [.01, .15], $p = .0207$, $n = 105$) had average weighted correlations that were positive and significantly different from zero. When setting HME calculations using latent factor scores as the reference group, pairwise t-test results showed that no significant differences in the average weighted correlation between the HME and children's math achievement were found. Thus, contrary to our expectations, HME scores calculated as latent factor scores did not result in significantly higher average weighted correlations between the HME and children's math achievement (sum scores: $b = 0.02$ [-0.01, 0.06], $t(2) = 1.28$, $p = .2012$; single items: $b = -0.06$ [-0.11, 0.004], $t(2) = -1.83$, $p = .0675$). The test for residual heterogeneity was significant ($Q_E[453] = 4279.26$, $p < .0001$), and even after accounting for the HME calculation used, $I^2 = 89.41\%$ of the variability in effect sizes was left unexplained.

Moderator Analyses for Math Assessment Characteristics

All moderation analysis results, including the effect sizes, sample sizes, and 95% confidence intervals for each math assessment moderator are presented in Figures 4 and 5. The results of the overall omnibus test with math assessment as the moderator showed that the math assessment used to measure children's math achievement was a significant source of heterogeneity [$F(13, 443) = 4.09$, $p < .0001$, $\sigma^2 = .04$ (.02, .07), $k = 456$] in the average weighted correlation between the HME and children's math achievement. Researcher-created assessments ($r = .14$ [.07, .20], $p < .0001$, $n = 188$), the KeyMath ($r = .11$ [.01, .20], $p = .0235$, $n = 13$), the WJ ($r = .48$ [.25, .71], $p < .0001$, $n = 3$), the WM ($r = .19$ [.11, .28], $p < .0001$, $n = 11$), the UENT-R ($r = .20$ [.09, .30], $p = .0003$, $n = 12$), the CATM ($r = .20$ [.07, .32], $p = .0026$, $n = 3$), the TEDI-MATH ($r = .16$ [.04, .27], $p = .0061$, $n = 11$), parent-report of children's math achievement ($r = .19$ [.09, .29], $p = .0001$, $n = 16$), multiple math assessments ($r = .16$ [.08, .24], $p = .0001$, $n = 29$), or the other math assessment category ($r = .13$ [.05, .20], $p = .0007$, $n = 95$)

had average weighted correlations that were positive and significantly different from zero, but the TEMA (r = -.09 [-.25, .08], p = .3156, n = 30), the PENS (r = .16 [-.13, .44], p = .2772, n = 41), or the CMA (r = .29 [-.10, .67], p = .1426, n = 4) did not. Pairwise t-test results showed that, when researcher-created math assessments were set as the reference group, the WJ (b = 0.34 [0.12, 0.57], t(12) = 3.01, p = .0028) had significantly higher average weighted correlations between the HME and children's math achievement, and the TEMA had significantly lower magnitude average weighted correlations (b = -0.22 [-0.40, -0.04], t(12) = -2.44, p = .0152). The test for residual heterogeneity was also significant (Q_E[443] = 4125.88, p < .0001), and even after accounting for the specific math assessment used, I^2 = 89.26% of the variability in effect sizes was left unexplained.

The results of overall omnibus test with math domain as the moderator showed that the math domain assessed was not a significant source of heterogeneity [F (4, 451) = 18.32, p = .0010, σ^2 = .04 (.02, .06), k = 456] in the average weighted correlation between the HME and children's math achievement. When arithmetic operations (r = .13 [.07, .19], p < .0001, n = 76), numerical relations (r = .09 [.03, .15], p = .0023, n = 70), numbering (r = .10 [.04, .16], p = .0016, n = 49), or multiple math domains (r = .15 [.09, .20], p < .0001, n = 257) were assessed the average weighted correlation between the HME and math achievement was positive and significantly different from zero, but when the spatial domain was assessed it was not (r =.00 [-0.14, 0.14], p = .9754, n = 4). According to pairwise t-test results with math measures assessing multiple math domains as the reference group, all other math domains (numerical relations: b = -0.05 [-0.08, -0.03], t(4) = -3.67, p = .0003; numbering: b = -0.05 [-0.08, -0.01], t(4) = -2.72, p = .0069; spatial skills: b = -0.15 [-0.28, -0.01], t(4) = -2.15, p = .0324), with the exception of arithmetic operations (b = -0.02 [-0.05, 0.01], t(4) = -1.20, p = .2325), had significantly lower average weighted correlations between the HME and children's math achievement. The test for residual heterogeneity was also significant (Q_E[451] = 4253.16, $p < .0001$), and even after accounting for the math domain assessed, I^2 = 89.40% of the variability in effect sizes was left unexplained.

For the symbolic math assessment moderator, the results of the overall omnibus test showed that whether math achievement was assessed using a symbolic, non-symbolic, or combination of symbolic and non-symbolic assessments was a significant source of heterogeneity [$F(3, 453) = 13.89$, $p < .0001$, $\sigma^2 = .04$ (.02, .06), $k = 456$] in the average weighted correlation between the HME and children's math achievement. Symbolic ($r = .15$ [.09, .21], $p < .0001$, $n = 213$), non-symbolic ($r = .12$ [.06, .18], $p < .0001$, $n = 61$), and combined symbolic and non-symbolic ($r = .13$ [0.07, 0.18], $p < .0001$, $n = 182$) math assessments all had average weighted correlations that were positive and significantly different from zero. According to pairwise t-tests with symbolic math assessments as the reference group, the average weighted correlation between the HME and children's math achievement was significantly lower when measured by a non-symbolic math assessment ($b = -0.03$ [-0.05, -0.01], $t(2) = -2.69$, $p = .0073$) or a combination of symbolic and non-symbolic math assessments ($b = -0.02$ [-0.03, -0.01], $t(2) = -3.68$, $p = .0003$). The test for residual heterogeneity was also significant (Q_E[453] = 4211.59, $p < .0001$), and even after accounting for whether the math assessment was symbolic, I^2 = 89.24% of the variability in effect sizes was left unexplained.

For the timed math assessment moderator, the results of the overall omnibus test showed that whether math achievement was assessed using a timed, untimed, or combination of timed and untimed assessments was a significant source of heterogeneity [$F(3, 453) = 14.38$, $p < .0001$, $\sigma^2 = .04$ (.02, .06), $k = 456$] in the average weighted correlation between the HME and children's math achievement. Math assessed by timed ($r = .07$ [.01, .13], $p = .0309$, $n = 67$), untimed ($r = .14$ [.09, .20], $p < .0001$, $n = 379$), or a combination of timed and untimed measures ($r = .14$ [.08, .21], $p < .0001$, $n = 10$) had average weighted correlations that were positive and significantly different from zero. In comparison to math assessed with a timed assessment, pairwise t -tests showed that math assessments that were untimed ($b = 0.07$ [0.04, 0.11], $t(2) = 3.89$, $p = .0001$) or a combination of timed and untimed ($b = 0.07$ [0.03, 0.12], $t(2) = 3.50$, $p = .0005$) showed significantly higher average weighted correlations between the HME and children's math achievement. The test for residual heterogeneity was also significant ($Q_E[453] = 4241.40$, $p < .0001$), and even after accounting for whether the math assessment was timed, $I^2 = 89.32\%$ of the variability in effect sizes was left unexplained.

For the composite math assessment moderator, the results of the overall omnibus test showed that whether math achievement was assessed using a composite or a single math assessment was a significant source of heterogeneity [$F(2, 454) = 12.28$, $p < .0001$, $\sigma^2 = .04$ (.02, .06), $k = 456$] in the average weighted correlation between the HME and children's math achievement. Both composite ($r = .15$ [.09, .22], $p < .0001$, $n = 70$) and single-measure ($r = .13$ [.08, .19], $p < .0001$, $n = 386$) math assessments had average weighted correlations that were positive and significantly different from zero. Pairwise t-test results showed that, in comparison to math assessed with a composite measure, the average weighted correlation between the HME and children's math achievement was statistically the same for math assessed with a single measure ($b = -0.02$ [-0.06, 0.02], $t(2) = -0.84$, $p = .3996$). The test for residual heterogeneity was also significant ($Q_E[454] = 4055.67$, $p < .0001$), and even after accounting for whether math was assessed using a composite or a single assessment, $I^2 = 88.81\%$ of the variability in effect sizes was left unexplained.

Finally, for the standardized math assessment moderator, the results of the overall omnibus test showed that whether math achievement was assessed using a standardized, unstandardized, or combination of standardized and unstandardized assessments was a significant source of heterogeneity [F (3, 453) = 9.62, p < .0001, σ^2 = .03 (.02, .06), k = 456] in the average weighted correlation between the HME and children's math achievement. Standardized (r = .15 [.09, .21], p < .0001, n = 85) and unstandardized math assessments (r = .12 [.07, .18], p < .0001, n = 357) had average weighted correlations that were positive and significantly different from zero, but combined standardized and unstandardized (r = .31 [-.05, .68], p = .1862, n = 14) math assessments did not. According to pairwise t-tests with standardized math assessments as the reference group, the average weighted correlation between the HME and children's math achievement was not significantly lower when measured by an unstandardized math assessment (b = -0.03 [-0.06, 0.00], t(2) = -1.79, p = .0744). The test for residual heterogeneity was also significant (Q_E[453] = 4015.51, p < .0001), and even after accounting for whether or not the study employed longitudinal or concurrent assessments, I^2 = 88.72% of the variability in effect sizes was left unexplained.

Moderator Analyses for Study Characteristics

All moderation analysis results, including the effect sizes, sample sizes, and 95% confidence intervals for the study characteristic moderator and its subgroups are presented in Figure 5. The results of the overall omnibus test with whether the study captured longitudinal (at different time points) or concurrent (at the same time point) relations between the HME and children's math achievement as the moderator showed that it was a significant source of heterogeneity [F (2, 454) = 14.93, p < .0001, σ^2 = .04 (.02, .06), k = 456] in the average weighted correlation between the HME and children's math achievement. Both longitudinal studies (r =

.09 [.02, .15], $p = .0105$, $n = 109$) and studies that measured the HME and children's math achievement concurrently ($r = .16$ [.10, .22], $p < .0001$, $n = 347$) had average weighted correlations that were positive and significantly different from zero. Pairwise t-tests showed that effect sizes that captured concurrent relations had average weighted correlations between the HME and children's math achievement that were significantly higher than effect sizes that captured longitudinal relations ($b = 0.07$ [0.02, 0.12], $t(1) = 2.55$, $p = .0112$). The test for residual heterogeneity was also significant ($Q_E[454] = 4279.81$, $p < .0001$), and even after accounting for whether the math assessment was timed, $I^2 = 89.39\%$ of the variability in effect sizes was left unexplained.

Overall, results from all 14 individual omnibus tests, except for the test including the HME component moderator, showed significant residual heterogeneity remained after accounting for the moderator modeled. This means that variability in the observed effect sizes was significantly larger than would be induced by sampling error alone, and that other moderators not tested in each moderator model were influencing the magnitude of the correlation between the HME and children's math achievement. Given that each moderator was tested individually, it is not surprising that no single moderator (with the exception of the HME component moderator) accounted for all significant effect size variance. Thus, as a final step to determine the amount of study heterogeneity accounted for by all potential moderators at once, a multilevel correlated effects meta-analysis was run with all 14 coded sample, assessment, and study features included. Results from the overall omnibus test that included all moderators showed that the combined moderators were a significant source of heterogeneity [$F(55, 333) = 4.08$, $p < .0001$, $\sigma^2 = .03$ (.01, .09)] in the average weighted correlation between the HME and children's math achievement. The residual heterogeneity of the model including all moderators was also significant ($Q_E(333) = 1028.12$, $p < .0001$, $k = 389$), with $I^2 = 67.61\%$ of the variability in effect sizes left unexplained, indicating that other moderators not tested in the present analysis were likely influencing the magnitude of the correlation between the HME and children's math achievement, beyond the effects of sampling error.

Publication Bias

Funnel plot. First, publication bias was assessed using a funnel plot of effect sizes (x-axis) to standard errors (y-axis), which is depicted in Figure 6. A visual inspection showed that most of the estimates, both below and above the mean, are clustered near the top of the funnel, suggesting high precision in effect size estimates overall. However, there are multiple studies outside of the shaded areas that represent the 90% (white), 95% (light grey), and 99% (dark grey) confidence intervals, suggesting that publication bias is likely. Based on Egger's test, which provides a parametric test for the skew of the distribution of effect sizes, significant publication bias is present ($t[454] = 2.19$, $p = .0289$), with slightly more effect sizes below, than above, the average weighted correlation coefficient. Given that slightly more lower magnitude correlations are reported than are higher magnitude correlations (i.e., above the meta-analytic average), our results do not support the existence of a file-drawer problem, wherein small effect sizes fail to be published and/or reported.

P-curve analysis. Results from the continuous p-curve analysis showed that both the full ($Z = -5.74$, $p < .0001$) and half ($Z = -5.51$, $p < .0001$) p-curve tests supported the existence of a significant right skew (see Figure 7). These combination test results, which have been shown to be more robust to p-hacking than a simple p-curve test (Simonsohn et al., 2014), indicated that the set of significant findings had evidential value. Furthermore, full p-curve, and both the half p-curve and binomial 33% power test were non-significant (full: $Z = 4.26$, $p > .9999$; half: $Z =$

4.58, $p > .9999$; binomial: $p > .9999$), indicating that the p-curve does not support that the evidential value is inadequate nor absent. These combined results indicate that the present meta-analytic sample of studies has evidential value and does not show evidence of *p*-hacking.

Sensitivity Analyses

Trim-and-fill. Trim-and-fill results are presented in Figure 8. The trim-and-fill procedure estimated that no studies were missing from above or below the 99% confidence interval around the average weighted correlation between the HME and children's math achievement, resulting in no studies being filled in. Even though a visual inspection of the funnel plot makes it appears as if some outliers exist outside the shaded confidence intervals, no studies were trimmed from the analysis, so our results do not support this. Given that no estimates were either trimmed or filled from our funnel plot, the same estimate of *r* based on the same number of studies ($n = 456$) was found for the trim-and-fill analysis as the overall meta-analytic results that did not utilize multilevel correlated effects modeling ($r = 0.08$ [0.07, 0.10], $p < .0001$). Overall, these results are promising because they indicate that research on the correlation between the HME and children's math achievement is likely not missing studies (i.e., does not have a file-drawer problem).

Fail-Safe N. According to the results of the fail-safe N test using the Rosenthal approach, in order to achieve null population results (i.e., $r = 0$), an additional 114,927 studies with null results (i.e., showing no significant association between the HME and children's math achievement) are needed to achieve the target null *p*-value of $> .05$. To achieve a $p > .01$, an additional 52,227 studies with null results ($r = 0$) are needed. These results show that our sample of effect sizes is likely capturing a true relation that is significantly different from zero.

Robust Variance Estimation. The results of the RVE analyses using robumeta indicated that the effect sizes, standard errors, and τ^2 values were identical across different values of ρ (r = .14, SE = .03, τ^2 = .02 for all values of ρ). Additionally, the overall estimate of the average weighted correlation between the HME and children's math achievement was the same as the meta-analytic results found using the metafor package of r = .14, indicating that our meta-analytic results using metafor are also robust.

Excluding a potentially influential study. Given that one study in our sample included over 227 effect sizes (Cheung, 2013), which may have exercised an inordinate influence on our overall results, we also conducted a sensitivity analysis to see if our results would change if the study was included in the average weighted correlation calculation. Results from the multilevel correlated effects meta-analysis that included Cheung, 2013 demonstrated an average weighted correlation between the HME and children's math achievement that was very similar to the average estimate that included all studies (Δ .01; r = .13 [.08, .19], p < .0001, σ^2 = .04 [.02, .06], k = 684). The model including the study also showed significant heterogeneity Q(683) = 4639.13, p < .0001, so moderator analyses were also performed to statistically test for the sources of between-study differences, with results reported in Appendix A.

Discussion

Given that there are individual differences in math achievement that appear prior to formal schooling that tend to persist once schooling begins, early intervention on math skill development is vital in order to combat the formation of math achievement gaps between children. One potential early influence on children's math achievement is the home math environment (HME), which captures a confluence of personal, environmental, social, and cultural factors that interact to impact children's math achievement. However, the association between the HME and children's math achievement has been found to vary widely between studies, with correlations ranging from small to large and positive to negative. Furthermore, there is little to no standardization in HME measurement across studies, with operationalizations of the HME including a range of different categories, like math-related activities (that directly and/or indirectly target math skills), parent math attitudes and/or beliefs, parent expectations for their children's math achievement, or any combination of the three. In addition, a significant body of research conducted on the association between parent math talk in the home and children's math achievement, has been previously excluded from traditional HME research area. Given the lack of consensus on the role of the HME in children's math achievement, the lack of consistency in how the HME is measured, and the previous separation between research on parent math talk and other more traditional HME components, we conducted the present meta-analysis in order to synthesize a previously disjointed research area and calculate the average weighted correlation between the HME and children's math achievement. Additionally, we conducted a series of moderator analyses in order to empirically test the impact of different sample, assessment, and study characteristics on the magnitude of the associations found.

Overall, the results of the present meta-analysis showed that the home math environment and children's math achievement had an average weighted correlation that is small and positive. However, given that the correlation of $r = .13$ translates to only 1% common variance between the two domains, it appears that, when evaluating across a combined sample that includes all empirical studies on the HME and children's math achievement, the overall role of the HME in children's math achievement is quite minimal. However, our moderator results did show some important subgroup differences, with higher magnitude correlations found when certain aspects of the HME, like parent math expectations ($r = .24$), or certain math assessments, like the WJ (r

= .48), were included in the effect size calculation. In fact, our moderation analysis revealed that a number of sample, assessment, and study characteristics were driving the variability in effect sizes found across studies.

Starting with the sample characteristic moderators tested, a developmental pattern emerged in which the association between the HME and children's math achievement diminished over time. Specifically, our results showed significant negative moderation by age, which was echoed by the significant moderation found for grade, with PK/KG grades demonstrating a higher correlation between the HME and children's math achievement than elementary grades. Given that once formal schooling begins, formal math instruction is likely to be more influential on children's math skill development than home environmental influences, this lower magnitude correlation between the HME and children's math achievement as children age or go from PK/KG to elementary school is not surprising. Indeed, a meta-analysis on the a home learning practices related to literacy and their relation with children's emergent literacy skills demonstrated the same decrease over time in the relation between shared book reading and children's literacy skills (Bus, Ijzendoorn, & Pellegrini, 1995).

The other sample characteristic that exercised a significant influence on the magnitude of the correlation between the HME and children's math achievement was the country of origin of the study sample. While most countries represented in our sample of studies aligned with our cumulative meta-analytic results, by demonstrating a small positive correlation between the HME and children's math achievement, samples from the Netherlands, which included 10 effect sizes, showed especially high correlations of $r = .49$. In contrast, U.S. samples, which represented the majority of our study sample, with 250 effect sizes, showed average weighted correlations below the cumulative average, with $r = .08$. These magnitude differences may be driven by differences in the educational systems of each country. Specifically, the Netherlands has a highly-stratified education system, which differentiates students according to tracks, with lower-performing students placed in vocational tracks, higher-performing students placed in professional tracks, and each track attending different schools with different curricula (Prokic-Breuer & Dronkers, 2012). Notably, the Netherlands has a larger vocational sector than other European countries with highly-stratified educational systems, creating a high selectivity for the professional track schools (Prokic-Breuer & Dronkers, 2012). This high level of competition to qualify for the professional track may be associated with more family involvement in children's education at home to try to prepare children for professional track schools, which translates into a higher magnitude correlation between the HME and children's math achievement. Contrastingly, the U.S. has a comprehensive educational system, in which both high- and low-performing students are taught in the same schools with undifferentiated curricula. Since children are not placed into schools based on performance, parents may not feel the same pressure to get their kids into the best schools, leading to more lax home learning practices. Importantly, this is only a potential explanation for the between-country differences found and was not directly tested in our moderation analyses. Notably, the Netherlands represents less than 5% of all effect sizes included in our meta-analytic sample, while the U.S. represents almost 40%, so the present results for moderation by country may not be comprehensive.

As a final note on the between-country differences, we were surprised to find that our hypothesis that Asian countries (i.e., China) would show significantly higher average weighted correlations between the HME and children's math achievement than European countries or the United States was not supported. Based on previous work showing that Chinese-American parents tend to teach math to their children in more formal, systematic ways than European-

American parents (Huntsinger et al., 2000), we expected the more dedicated focus on math instruction to translate into higher magnitude correlations between the HME and children's math achievement for Chinese samples, especially since direct math activities typically demonstrate higher magnitude correlations with children's math achievement than indirect activities (cite). However, Chinese samples did not show a significant association between the HME and children's math achievement at all. Furthermore, our HME component moderation analyses showed that direct activities were no more related to children's math achievement than indirect activities. The lack of association found for Chinese samples may be partially explained by the fact that China was the only Asian country represented in our study sample, and that most of the effect sizes for Chinese samples came from one study (Cheung, 2017). This points to the need for more diversity in HME research to capture the true impact of cultural differences in parenting practices on the home learning environment and their link to children's achievement outcomes.

Our hypotheses for our special sample characteristic moderators were also not supported. Specifically, no magnitude differences favoring high-SES samples and boys were found between low- and high-SES samples or boys and girls. Our findings for SES were unexpected, given that low SES environments have been shown to be less favorable overall with less resource availability, in terms of tangible items, like books and flashcards, and nontangible items, like parent time and attention (Duncan & Magnuson, 2005). We would have expected this resource deficit to manifest as a lower quality HME, resulting in a lower magnitude correlation between the HME and children's math achievement for low SES samples, but this was not the case. Since our samples were limited, we would have benefitted from having more studies with high- and/or low-SES samples to make more definitive conclusions about the role of SES in the association between the HME and children's math achievement.

For boys and girls, the findings on math achievement differences have been mixed, with meta-analytic work showing that math performance is similar between genders, but that boys show more positive math attitudes and affect than girls (Else-Quest, Hyde, & Linn, 2010). Given that Eccles' Parent Socialization Theory conjectures that parent math attitudes are social reinforcers for children's beliefs about their math abilities and math in general, also impacting children's math performance (Eccles et al., 1990), we expected parent socialization differences for female versus male children captured by the HME, to result in different magnitude correlations for all-boy versus all-girl samples. Differences could have manifested in either direction, with the HME having a more pronounced impact on girls' performance because it is where their less positive attitudes and beliefs toward math originated, or with the HME having a lesser impact because their more negative attitudes toward math would thwart the potential positive impact of the HME on their math achievement. However, this was not found to be the case across the present study sample, as boys and girls had correlations that were statistically the same. Overall, this finding may be positive because it indicates that qualitative differences in the HME that parents provide for children are probably not found based on child gender.

One final notable finding for special sample characteristics was that SLI samples demonstrated especially high correlations. Since only one study and 4 effect sizes included a sample composed of 33% SLI children (Kleemans et al., 2013), we caution over drawing any over-arching conclusions based on this special sample characteristic. However, given that the study sample also comes from the Netherlands, which demonstrated the highest magnitude correlation of any other country, it may be the case that the high magnitude correlations found for SLI samples are actually attributable to the sample's country of origin. We are unable to parse these relations based on the analyses run in the present study, so in line with recent work on math achievement (Wei, Lenz, & Blackorby, 2013), future work should examine whether the association between the HME and children's math achievement differs in samples with learning-related difficulties.

Across all assessment moderators tested, results revealed that differences in how the HME and math achievement were measured significantly impacted the magnitude of the correlation between the HME and children's math achievement. Overall, this is an important finding because inconsistent measurement methods, utilizing a wide array of measurement instruments for both the HME and math, are common in HME research. Thus, it makes sense that studies within the same research area have reflected such disparate findings due to measurement inconsistency. Focusing first on how the nuances in HME assessment impacted how closely the HME was associated with children's math achievement, we found that contrary to our hypothesis, there were no significant differences found for correlations that included HME measures of direct activities (i.e., math-related activities that directly targeted math skills) versus indirect activities (i.e., math-related activities that incidentally targeted math skills; LeFevre et al., 2009; Skwarchuk et al., 2014). On the other hand, our hypothesis that parent attitudes and/or beliefs toward math would have lower magnitude correlations with children's math achievement than direct activities was partially supported, with parent attitudes and/or beliefs toward math demonstrating lower magnitude correlations than HME measures that included a combination of direct and indirect activities or only indirect activities. Our hypothesis that parent expectations for children's math achievement would have the highest magnitude correlations of all HME components was also partially supported. Specifically, parent expectations for their children's math achievement had significantly higher magnitude correlations with children's math achievement than HME measures that included a combination of direct and indirect math activities or spatial activities.

Surprisingly, the correlations that included parent attitudes and/or beliefs toward math were not found to be any weaker than the correlations that included parent expectations for their children's math achievement. This violated our hypothesis, based on meta-analytic work on the association between parent involvement and children's achievement outcomes, that parent math expectations would be the most influential facet of the HME overall and the most important social influence in the HME. Although parent math expectations were found to matter more than math-related activities, our results showed that any parent social influences related to math, both in terms of affective factors and expectations, were equally important for children's math performance. Overall, these nuanced differences driven by HME measurement inconsistency point to the need for the creation of a standardized measure of the HME that can be used across different studies and samples so that differences found between studies may reflect true differences, rather than measurement differences.

Looking specifically at how children's math skills were assessed both the math assessment used and the math domain measured were found to be significant moderators. In comparison to researcher-created math assessments, only the WJ was more closely linked to the HME. Importantly, this difference is likely not explained by the fact that the WJ is a standardized assessment, while researcher-created assessments are unstandardized, because our moderator results showed that the correlations between the HME and children's math achievement were statistically the same for standardized versus unstandardized math assessments. Relatedly, the differences found for the WJ could be attributable to the fact that only two studies and three effect sizes included the WJ. Although the two studies had many differences (Niklas et al., 2016;

Skwarchuk, 2009), like samples from different countries, and HME assessment of different components, one feature they had in common was their inclusion of multiple math domains in their effect sizes. However, given that the average weighted correlation for math assessed by multiple math domains of r = .14 is very similar to the average weighted correlation found without accounting for moderators (r = .13), and math assessed by the WJ had a significantly higher magnitude correlation of r = .48, the measurement of multiple math domains is probably not driving these results. It may be the case that there is something about the WJ, other than being standardized, and including subtests that assess multiple math domains, that is driving the large effect sizes found, but the reasons cannot be clarified by the present analyses.

As previously mentioned, when looking at the moderating effect of math domain, our results revealed that math assessments measuring multiple math domains showed the highest magnitude correlations between the HME and children's math achievement compared to all other math domains, except arithmetic operations. The finding that multiple math domains result in stronger effects is not surprising due to the multi-faceted nature of math ability in children, which is made up of many underlying skills (Geary, 2004). Accordingly, any assessment of math that captures the broad range of skills that underlie math ability is likely to provide a more complete measure of math ability overall (Purpura & Lonigan, 2013). Thus, future HME work would do well to measure multiple math domains at once in order to capture the role of the HME in children's math achievement more fully.

Another important finding on the moderating effects of math domain was that math assessed as arithmetic operations was more strongly associated with the HME than math assessed in the domain of numerical relations. Some previous work suggests that both numbering and numerical relation skills are needed in order for arithmetic operations skills to develop

(Aunio & Niemivirta, 2010), which aligns with the perspective that children's early math skills are cumulative and build on each other, becoming increasingly complex over time (Sarama & Clements, 2009). Thus, the fact that the HME emerges as a stronger influence in complex math domains that require children to draw on pre-existing foundational math skills may be an indication that the benefit of the HME is that it gives children a foundation of math skills that can help them build up more complex math knowledge and understanding.

Our moderator analyses also highlighted the importance of methodological differences in HME-math achievement studies, as many math assessment and study features were found to influence correlation magnitudes. In fact, all math assessment features tested were found to be significant sources of heterogeneity, including whether math was measured using a standardized or unstandardized assessment, a composite or single assessment, a timed or untimed assessment, or a symbolic or non-symbolic assessment. Alongside our findings of significant moderation by math assessment and math domain, these results mean that caution should be taken when choosing math assessments to investigate the link between the HME and children's math achievement, as various assessment features may drive variability in results. Therefore, the best approach to take when investigating the HME-math achievement link may be to include multiple math assessments in order capture the wide array of possible math assessment differences. One final moderator result worth noting was that studies that reported concurrent correlations were found to be stronger than studies that reported longitudinal correlations. Specifically, when the HME and children's math achievement were measured at the same time point, a higher magnitude correlation was found, whereas assessments of the HME and children's math achievement that occurred at different times points had lower magnitude correlations. This also points to the need for caution in HME study design, as measurement timing is likely to play a part in the consistency of study results. Thus, a study that can incorporate both concurrent and longitudinal relations between the HME and children's math achievement would be ideal.

Overall, the present meta-analysis provides a robust estimate of the true correlation between the HME and children's math achievement. Although our tests of publication bias were significant, our p-curve analysis confirmed that the results in our study sample had evidential value. Furthermore, our sensitivity analyses indicated that our findings were almost identical across different values of ρ and if a study reporting almost 30% of all effect sizes (Cheung, 2013) was removed. Based on these statistical tests and sensitivity analyses, the meta-analytic results reported here most likely represent true, robust effects.

Conclusion

In general, the small, positive average weighted correlation of $r = .13$ found between the HME and children's math achievement indicates that although the HME is significantly associated with children's math achievement, with a higher degree of parent-child math interaction and socialization associated with better child math performance, the association is limited. One reason behind the small HME-math achievement link may be the widespread use of surveys to measure the HME. There may be important aspects of the HME, like the quality of math-related interactions instead of just their frequency, which surveys cannot properly capture, for which direct observation techniques would be better suited. Although the low magnitude correlation found for parent math talk, which is typically measured through direct observation, may seem to contradict this proposition, it may be the actual HME component of parent math-related utterances that drives the lower magnitude correlations between parent math talk and children's math achievement, rather than the observational measurement techniques used. Thus, further measurement work based on sound psychometric theory in order to create and pilot observational HME instruments is needed in future HME research. In addition, based on moderator results demonstrating that measurement inconsistencies in how the HME and math achievement are assessed are largely driving study heterogeneity, future HME work that employs survey-based measurement should aim to include as many facets of the HME and math achievement as possible in order to get a more complete picture of the true relations.

Although the meta-analytic results here did not demonstrate a strong relation between the HME and children's math achievement, the fact that a significant correlation was found between the two domains provides evidence to support the importance of home-based math learning. Although other individual, cultural, social, and environmental factors may create differences in the HME that impact its association with children's math achievement, the HME was found to be a positive influence on children's math achievement, overall. Importantly, more empirical work needs to be done to implement consistent and comprehensive measurement of the HME, in order to fully understand its role in children's math achievement. Given that our analysis revealed that there are unmeasured characteristics driving our results, it is important to conduct more research on the HME and children's math achievement using a battery of HME and math assessments, while including diverse samples across a range of countries in order to determine how the HME can be utilized for early math intervention.

Table 1. Article coding key

Category	Value Description
Sample Characteristics	
Grade	1 = PK/KG only
	2 = Combination PK/KG and elementary school
	3 = Elementary school only
	4 = Combination elementary and middle school
	5 = Middle school only
Country	1 = United States
	2 = Canada
	3 = Netherlands
	5 = Germany
	6 = Greece
	7 = Italy
	8 = China
	9 = Chile
	11 = Australia
	10 = Other
Special Sample Characteristics	1 = None, average sample
	2 = Low SES
	3 = High minority (30% or more)
	4 = All one ethnicity (75% or more)
	5 = All girls
	6 = All girls
	7 = SLI sample (30% or more)
	8 = Other
	9 = High SES
HME Assessment Characteristics	
HME Component	2 = Direct activities
	3 = Indirect activities
	4 = Attitudes and/or beliefs
	5 = Math expectations
	6 = Spatial activities
	7 = Math talk
	9 = Combination of direct and indirect activities
	11 = Combination of activities and attitudes
HME Calculation	1 = Latent factor score
	2 = Sum score
	3 = Single item
Math Assessment Characteristics	
Math Assessment	1 = Researcher-created
	2 = KeyMath
	4 = TEMA
	5 = WJ
	6 = WM

Table 1. Article coding key – continued

Category	Value Description
Math Assessment – continued	7 = UENT-R
	8 = PENS
	9 = CATM
	11 = CMA
	18 = TEDI-MATH
	13 = Multiple math assessments
	12 = Other
Math Domain	2 = Arithmetic operations
	3 = Numerical relations
	4 = Numbering
	10 = Multiple math domains
	11 = Spatial skills
Symbolic	1 = Symbolic
	2 = Non-symbolic
	3 = Combination of both
Timed	1 = Timed
	2 = Untimed
	3 = Combination of both
Composite	1 = Composite
	2 = Single Measure
Standardized	1 = Standardized
	2 = Unstandardized
	3 = Combination of both
Study Characteristics	
Longitudinal	1 = Longitudinal relation
	2 = Concurrent relation

Note. For some moderators, numbering is out of order with missing values due to changes to the coding scheme; TEMA = Test of Early Mathematics Ability; WJ = Woodcock Johnson-III Tests of Achievement; WM = Woodcock–Muñoz Batería III—Spanish adaptation of the Woodcock–Johnson Tests of Achievement; Utrecht Early Numeracy Test- Revised; PENS = Preschool Early Numeracy Skills test; CMA = Child Math Assessment; SLI = Specific Language Impairment

Table 2. Math assessment descriptions

Assessment	Description	Reliability (α)
Key Math-3 Diagnostic Assessment (KeyMath)	The KeyMath-3[1] is comprised of ten subtests, including Numeration, Algebra, Geometry, Measurement, Data Analysis, and Probability, which measure basic math concepts, Mental Computation and Estimation, Addition and Subtraction, Multiplication and Division, which measure math operations, and Foundations of Problem Solving and Applied Problem Solving, which measure math applications, all based on the National Council of Teachers of Mathematics Principles and Standards for School Mathematics. The tests are untimed, norm-referenced, and available for ages 4 years, 6 months through 21in alternate forms A and B to assesses key mathematical concepts and skills.[1,2]	.86 - .99[1]
Test of Early Mathematics Ability (TEMA), 2 and 3	The TEMA[3] is a norm-referenced or diagnostic test used to identify the level at which children ages 3 to 8.11 years are performing in a variety of specific mathematics skills based on the National Council of Teachers of Mathematics curriculum requirements.[4,5] The assessment takes 40 minutes and is available in alternate forms A and B.4	.92[5]
Woodcock-Johnson III/IV Tests of Achievement (WJ)	The WJ[6] is a norm-referenced assessment, which includes four subtests (Calculation, Math Fluency, Applied Problems, and Quantitative Concepts) that measure numeracy, arithmetic calculation, and math reasoning ability. The assessment items are presented both visually and orally and progressively increase in difficulty. Alternate forms A and B are available for every subtest.[6,7]	.85 - .92[6]
Woodcock-Muñoz Batería III (WM)	The WM[8] is the Spanish-language version of the WJ, which has all the same features and subtests as the WJ but was developed with native Spanish speakers.[8,9]	.90[8]
Utrecht Early Numeracy Test – Revised (UENT-R)	The UENT-R[10,11] is comprised of 40 items, which measure numerous foundational mathematical concepts, namely comparison, classification, one-to-one correspondence, counting, seriation, and general number knowledge. The test is given one on one, with alternate forms A and B available and must be completed within 30 minutes.[12]	.84 - .90[12]

Table 2. Math assessment descriptions – continued

Assessment	Description	Reliability (α)
Preschool Early Numeracy Skills test (PENS)	The PENS[13] is a researcher-created math assessment comprised of three subtests, including Numbering, Numerical Relations, and Arithmetic Operations, which measure the breadth and depth of children's numeracy competence from ages 3 to 5. The unstandardized, untimed test is used as a progress-monitoring assessment or screener and measures children's knowledge of symbolic numbers and mathematical terms and concepts with 24 total items of increasing difficulty.[13,14]	
Deutscher Mathematiktest für erste Klassen ("German Mathematical Test for first graders"; DEMAT 1+)	The DEMAT 1+[15] is a German-language, curriculum-based, standardized math assessment that includes nine subtests, which measure children's mathematical abilities, like arithmetic calculation and symbolic number knowledge, in order to pinpoint children's math-related strengths and weaknesses. The DEMAT exists in different forms for grades 1 through 4 and is administered as a group test with a 40-minute limit or as a single test with a limit of 20-35 minutes. Alternate forms A and B available for each grade level.[15,16]	.89[16]
Child Math Assessment (CMA)	The CMA[17] is a researcher-developed math assessment based on the standards set by the National Council of Teachers of Mathematics (2000), which measures 3- to 5-year-old children's informal mathematical knowledge. The CMA is administered individually and is comprised of fifteen tasks that cover five math domains: number sense, arithmetic, geometric reasoning, pattern knowledge, and measurement.[18]	.90[18]
Individual Growth & Development Indicators of Early Numeracy (IGDIs-EN)	The IGDIs-EN[19] is a standardized, curriculum-based math assessment for 3- to 5-year old children that is administered individually for use in universal screening and progress monitoring of children's early math skills. The subtests include one timed assessment—oral counting (number of items varies)—and three untimed assessments: number naming (63 items), quantity comparison (32 items), and one-to-one correspondence counting (20 items).[19,20]	.62-.91 (test-retest)[20]

Table 2. Math assessment descriptions – continued

Assessment	Description	Reliability (α)
Bracken Basic Concepts Scale- 3rd Edition: Receptive (BBCS:3-R) and School Readiness Composite (BBCS:3-SRC)	The BBCS:3-R and BBCS:3-SRC[21] are standardized assessments for children ages 3 to 6 years 11 months, which non-verbally measure children's basic math, cognitive, and language development. There are ten assessment domains, which measure color, shape, size, number, and other concepts not related to math. Each subtest is comprised of 10-22 questions that measure children's receptive knowledge by children to indicate their response to each item by pointing to the correct answer.[20]	.78-.97[21]
Performance Indicators in Primary School (PIPS)	The PIPS[22] is a standardized assessment that is group-administered at beginning of primary school (4-5 years old), and subsequently, at the end of the first (5-6 years old), third (7-8 years old), fifth (9-10 years old), and seventh grades (11 years old). It measures children's reading, math, and social development. The math assessments measure basic knowledge of math language, counting, simple arithmetic, number and shape recognition, and more advanced mathematical procedures as children advance in school. Different test items of increasing difficulty are administered to children as they progress in age/grade.[22,23]	.92[24]
Early Childhood Longitudinal Study- Kindergarten Cohort Math assessment (ECLS-K Math)	The ECLS-K[25] math assessments were administered in kindergarten (fall and spring) and first grade to measure children's conceptual and procedural math knowledge and problem-solving ability based on national and state standards for mathetmatics development. All assessments were untimed and given orally and individually based on a standardized protocol.[26]	.92-.94[26]
The Research-Based Early Mathematics Assessment Short-Form (REMA-S)	The REMA-S[27] is a researcher-created, curriculum-based math assessment that measures 3- to 5-year-old children's early math knowledge in the domains of numbers (recognition, comparison, sequencing, counting), arithmetic, and geometry. The REMA-S is untimed and includes 19 total items, given orally by a trained test administrator and is based on learning trajectories.[27,28]	1.00 (item reliability)[28]

Table 2. Math assessment descriptions – continued

Assessment	Description	Reliability (α)
Test for Diagnostic Assessment of Mathematical Disabilities (TEDI-MATH)	TEDI-MATH[29] is a standardized assessment designed for diagnosing math-related disorders in children up to third grade (4-9 years old). TEDI-MATH includes six subtests that measure four math-related domains: number logic (classification and seriation), counting, numerosity, and symbolic number knowledge (including arithmetic computation). The test is untimed and administered individually.[29,30]	.70-.97[29]
California Achievement Test Mathematics subtest (CATM)	The CAT[31] is a normed, standardized test administered to students individually, which assesses reading, language, and math ability for grades K-12, using both multiple choice and short answer questions (Tiegs & Clark, 1977). The CATM is comprised of six mathematical subtests and 92 items, which are based on state and district curricula, and measure children's computation skills and concepts and applications (numeration, problem-solving, measurement, and geometry).[31,32]	.98[32]
Diagnostic Test for Basic Mathematical Concepts (DTBMC)	The DTBMC[33] is a standardized, curriculum-based math assessment comprised of five subtests, which measure basic math knowledge of ordinal numbers, cardinal numbers and basic math concepts, number identification, word problems, and basic arithmetic, with tasks that include items of increasing difficulty. The DTBMC can be administered individually or in a group setting and is not timed.[33,34]	.94 (test-retest)[35]
Stanford Diagnostic Mathematics Test, Fourth Edition (SDMT4) Computation subtest	The SDMT[36] is a nationally-normed test available in different forms for grades 2 through 12, which includes two subtests (Concepts and Applications, Computation) that measure children's knowledge of the basic math concepts and skills that are foundational to the development of problem-solving ability.[37] The Computation subtest includes 20 grade-level questions that require the application of addition and subtraction procedures with a 25-minute time-limit for completion.[38]	.85[38]

Note. Citations: [1]Connolly, 2007; [2]Skwarchuk, Sowinski, & LeFevre, 2014; [3]Ginsburg & Baroody, 2003; [4]https://www.parinc.com/Products/Pkey/442; [5]Methe, Hintze, & Floyd, 2008; [6]Woodcock, McGrew, & Mather, 2001; [7]Skwarchuk, 2009; [8]Muñoz-Sandoval, Woodcock, McGrew, & Mather, 2005; [9]del Rio, Susspereguy, Strasser, & Salinas, 2017;[10]Van Luit, Van de Rijt and Pennings, 1994; [11]Van Luit & Van de Rijt, 2009; [12]Passolunghi, Lanfranchi, Altoè, &

Sollazzo, 2015; [13]Purpura, 2009; [14]Purpura, Reid, Eiland, & Baroody, 2015; [15]Krajewski, Küspert, Schneider, & Visé, 2002; [16]Niklas & Schneider, 2014; [17]Klein, Starkey, & Wakely, 2000; [18]Klein, Starkey, Clements, Sarama, & Iyer, 2008; [19]Hojnoski & Floyd, 2013; [20]Missall, Hojnoski, Caskie, & Repasky, 2015; [21]Bracken, 2006; [22]Tymms & Albone, 2002; [23]Bull, Espy, & Wiebe, 2008; [24]Tymms, Merrell, Hawker, & Nicholson, 2014; [25]NCES, 2002; [26]Puccioni, 2015; [27]Weiland et al., 2012; [28]Zippert & Rittle-Johnson, 2018; [29]Gregoire, Noël, & Van Nieuwenhoven, 2004; [30]Yildiz, Sasanguie, De Smedt, & Reynovet, 2018; [31]Tiegs & Clark, 1977; [32]Mboya, 1986; [33]Ikäheimo, 1996; [34]Aunola, Leskinen, Lerkkanen, & Nurmi, 2004; [35]Silinskas, Leppanen, Aunola, Parrila, & Nurmi, 2010; [36] Lichtenberger, 2008; [37]Wang, 2004; [38]Vukovic, Roberts, & Wright, 2013.

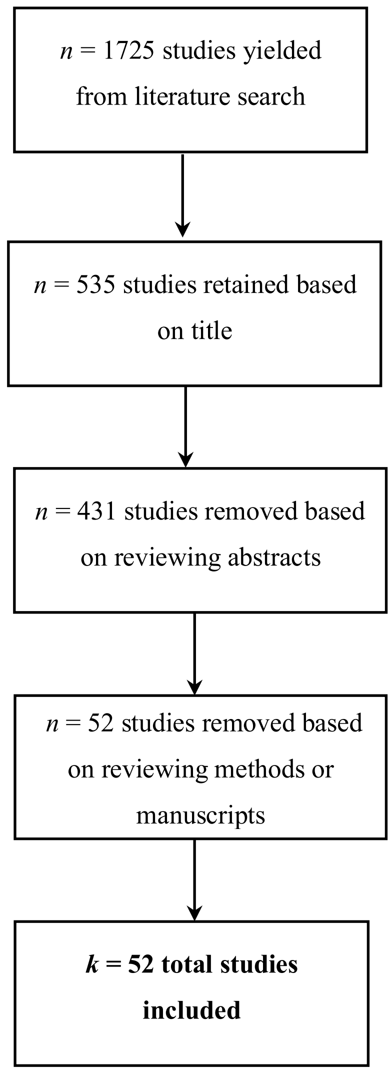

Figure 1. Article selection flow chart.

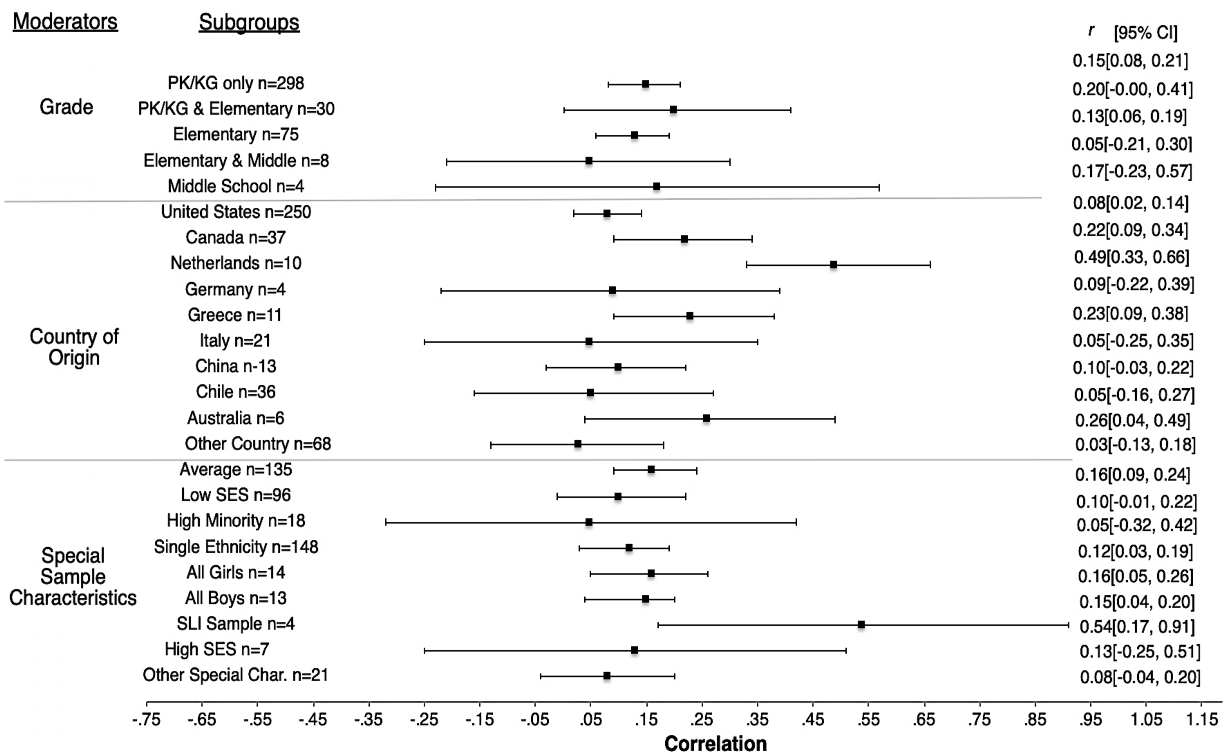

Figure 2. Effect size estimates (x-axis) for sample characteristic moderators, including grade, country of origin, and special sample characteristics with the number of effect sizes (*n*) for each moderator subgroup (left column) and the correlation coefficient (*r*) and its corresponding 95% confidence interval (far right column).

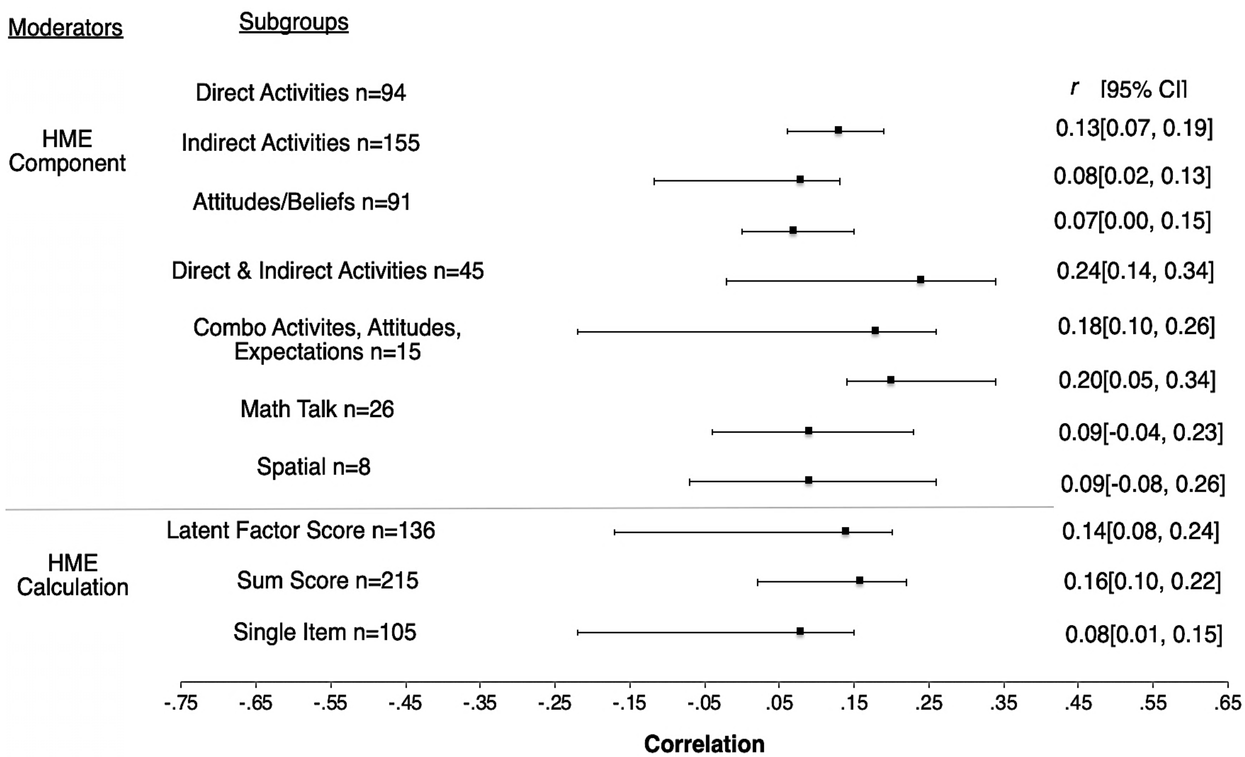

Figure 3. Effect size estimates (x-axis) for HME assessment moderators, including HME component and HME calculation with the number of effect sizes (n) for each moderator subgroup (left column) and the correlation coefficient (r) and its corresponding 95% confidence interval (far right column).

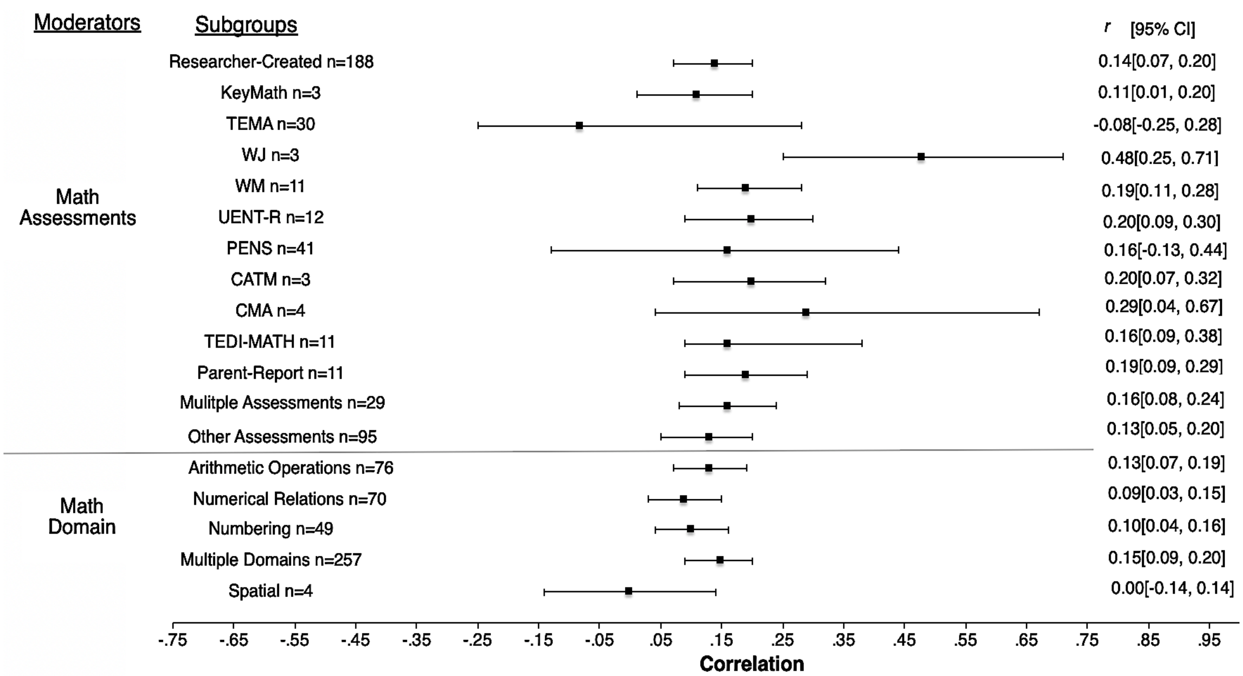

Figure 4. Effect size estimates (x-axis) for math assessment moderators, including math assessment and math domain with the number of effect sizes (*n*) for each moderator subgroup (left column) and the correlation coefficient (*r*) and its corresponding 95% confidence interval (far right column).

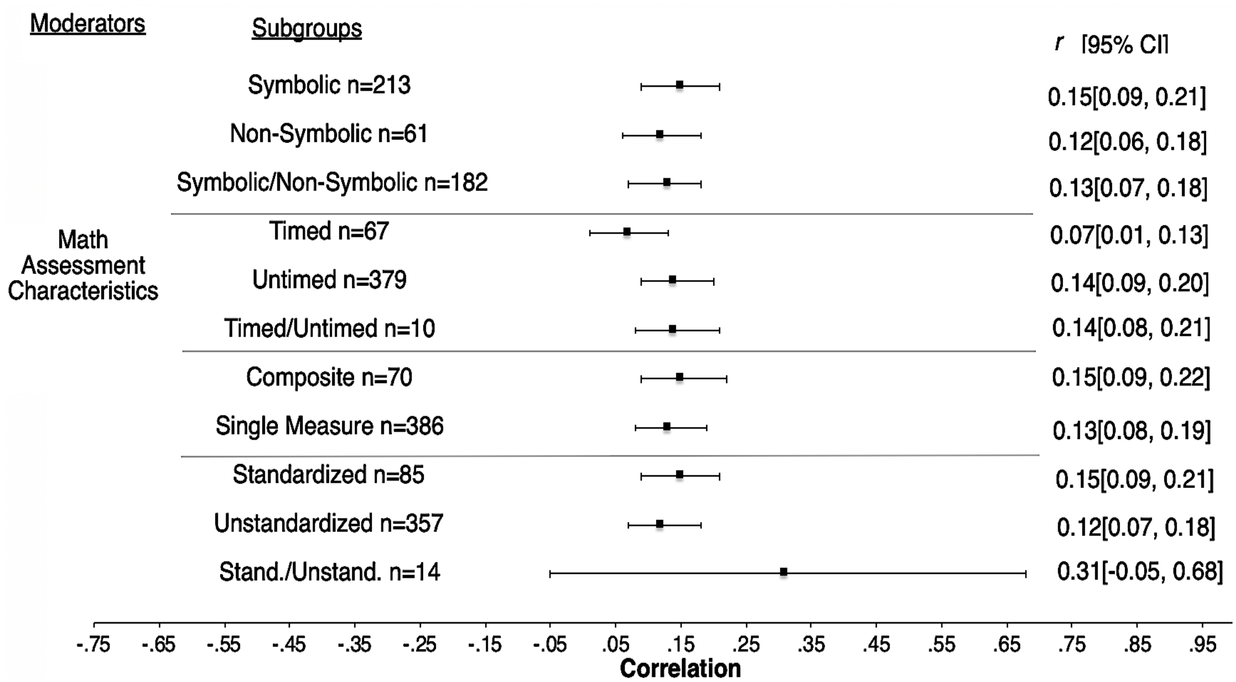

Figure 5. Effect size estimates (x-axis) for math assessment moderators, including whether the assessment was symbolic, timed, a composite, or standardized, with the number of effect sizes (*n*) for each moderator subgroup (left column) and the correlation coefficient (*r*) and its corresponding 95% confidence interval (far right column).

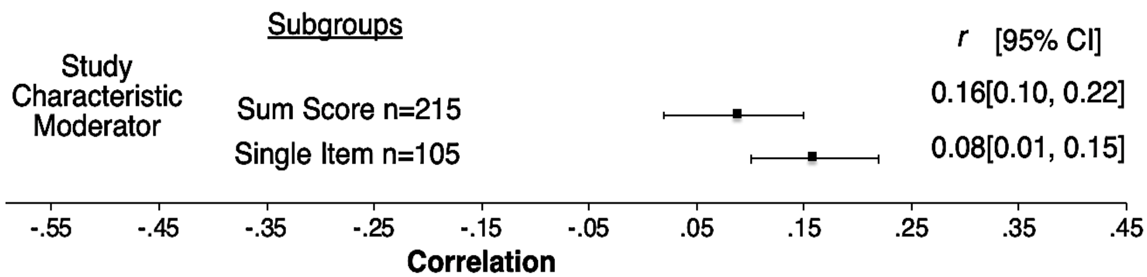

Figure 6. Effect size estimates (x-axis) for the study moderator of whether the study reported longitudinal or concurrent effect sizes with the number of effect sizes (*n*) for each moderator subgroup (left column) and the correlation coefficient (*r*) and its corresponding 95% confidence interval (far right column).

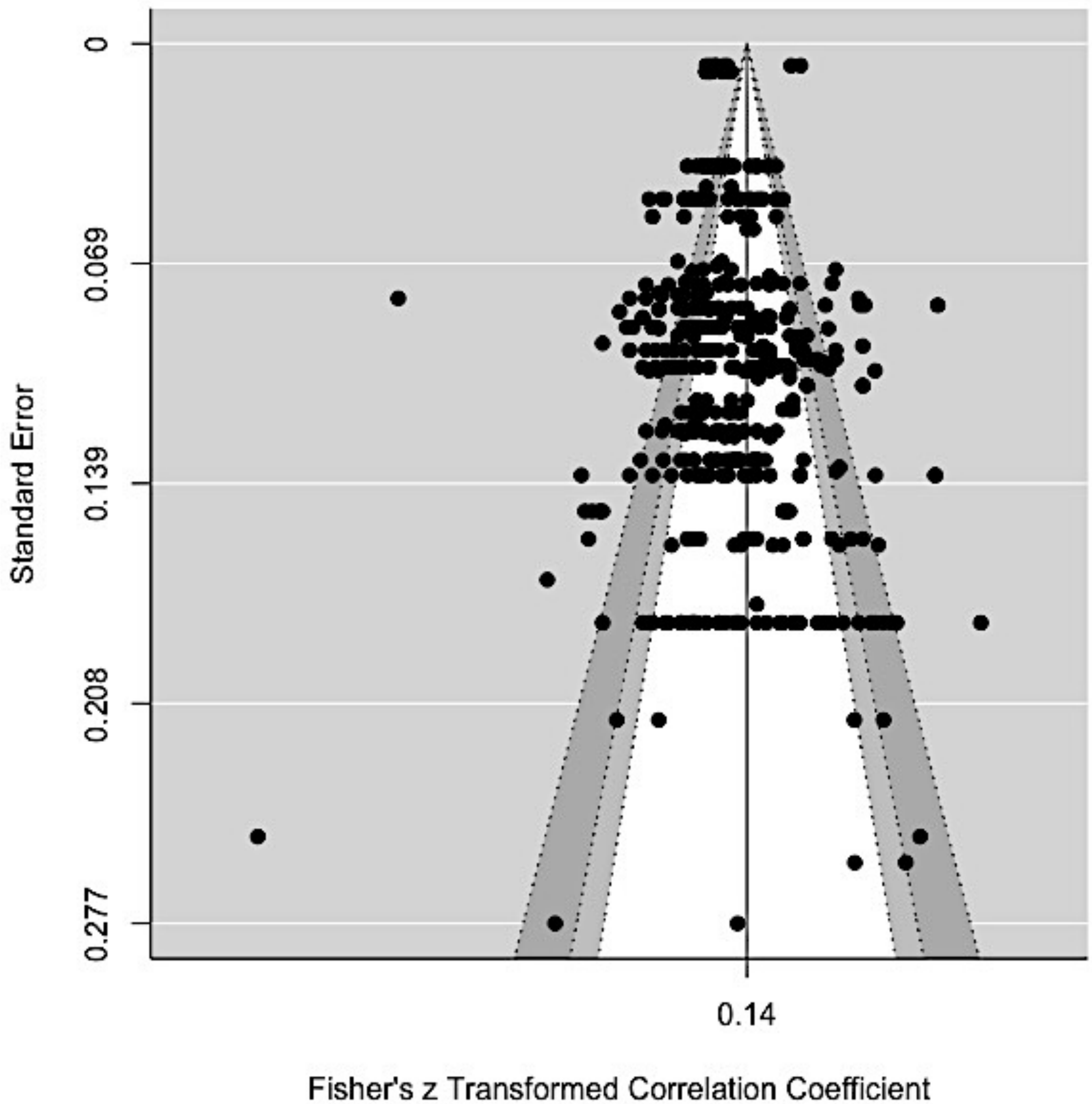

Figure 7. Enhanced funnel plot of the multilevel correlated effects meta-analysis results with the average weighted correlation of r = .14 and confidence intervals on the 90^{th} (white), 95^{th} (gray), and 99^{th} (dark gray) percentiles.

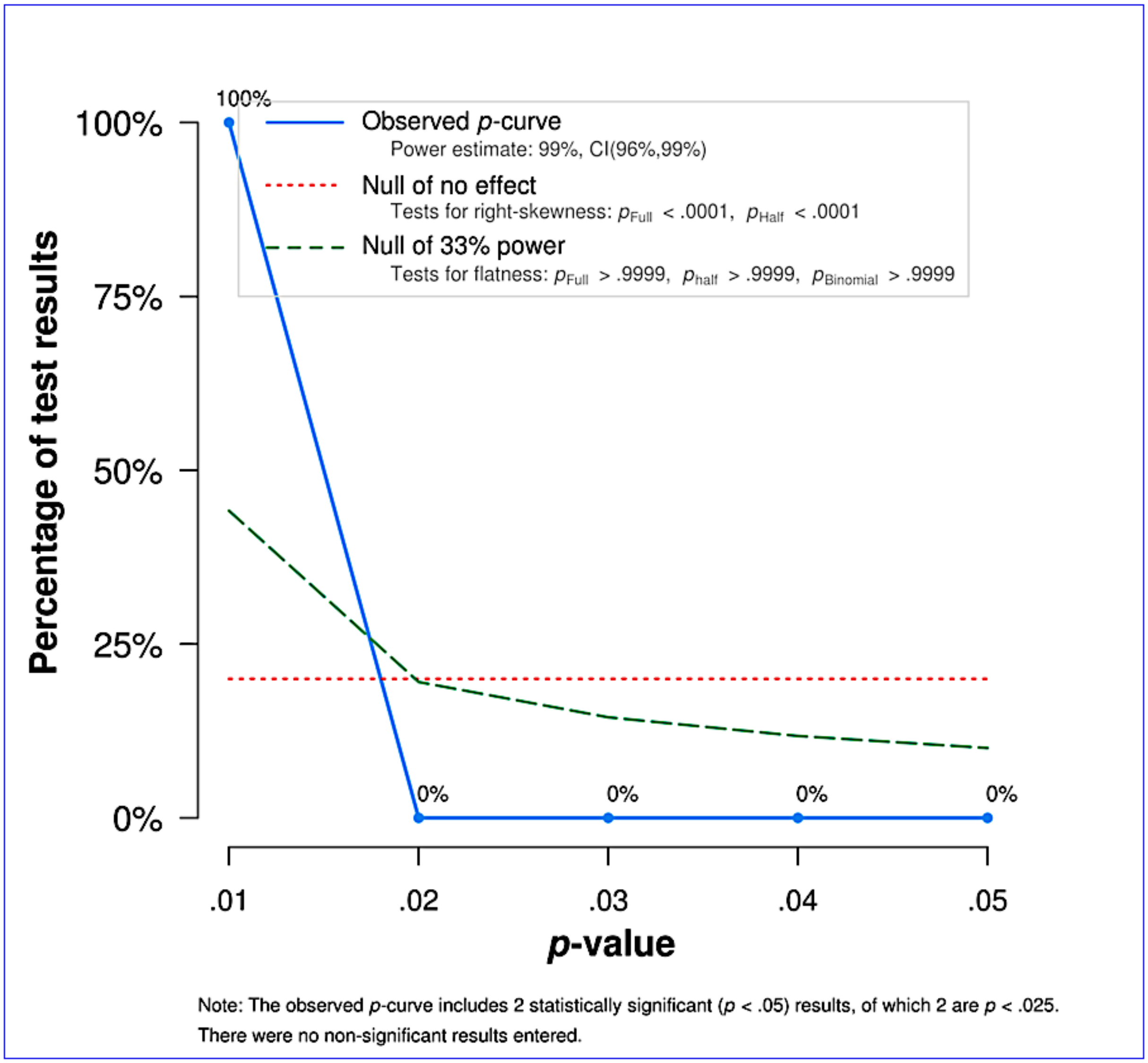

Figure 8. P-curve analysis results showing that a significant right skew and non-significant results at 33% power for both the full and half p-curve tests, indicating evidential value and a lack of *p*-hacking in our study sample.

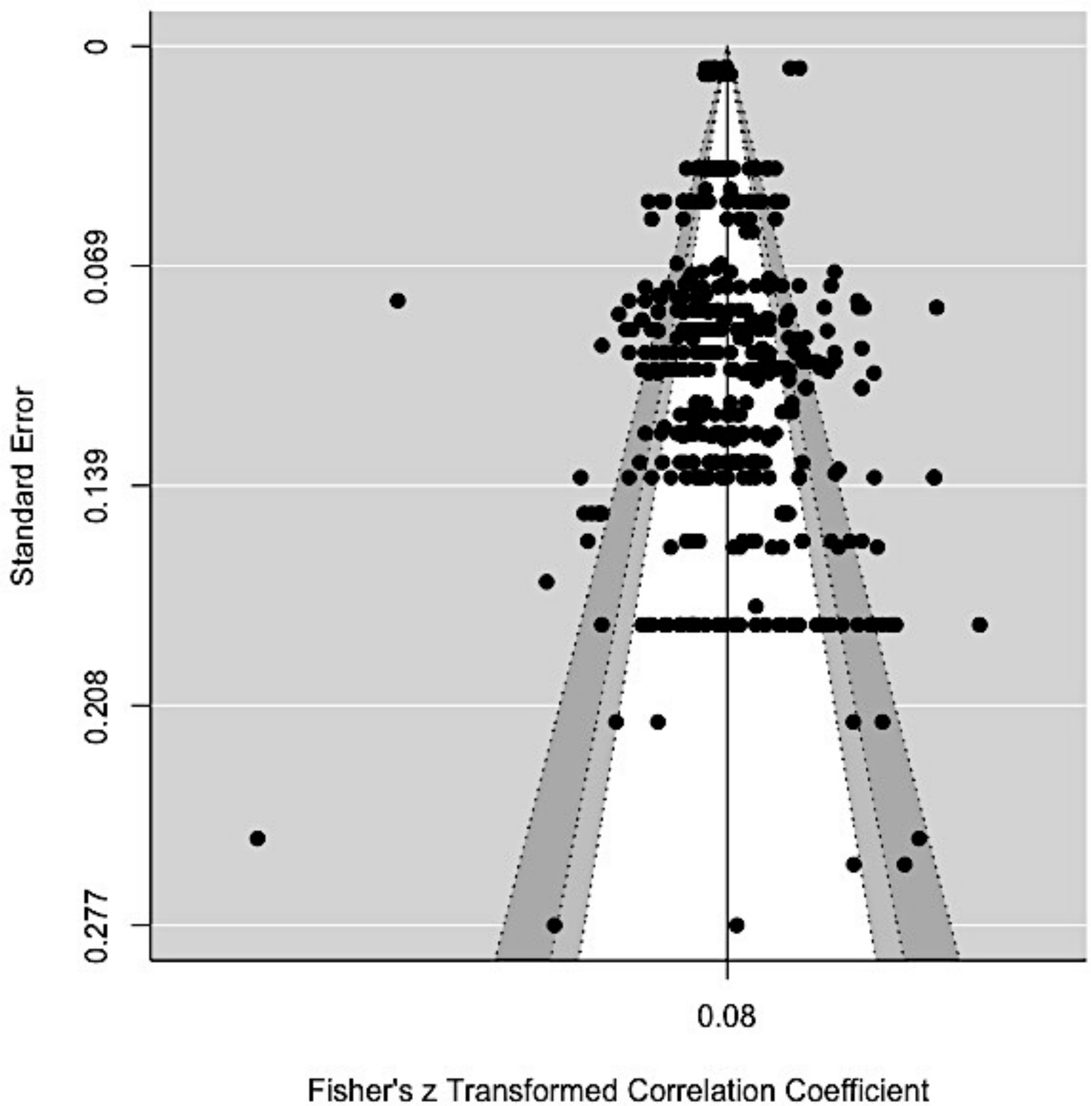

Figure 9. Trim-and-fill plot of the random effects meta-analysis average weighted correlation of $r = .08$ and confidence intervals on the 90th (white), 95th (gray), and 99th (dark gray) percentiles. The trim-and-fill method does not allow for the inclusion of a multilevel object, so the average weighted correlation coefficient that did not utilize multilevel correlated effects modeling was set as the mean effect size ($r = .08$).

APPENDIX A

RESULTS INCLUDING CHEUNG, 2013

Overall Average Weighted Correlation Between the Home Math Environment and Children's Math Achievement

Results of the random-effects analysis, which did not account for study dependence, yielded an average weighted correlation of .07 [.05-.08], *SE*=.01, $p < .0001$. However, given that only 52 studies resulted in a sample of 684 effect sizes, a multilevel correlated effects analysis was conducted to account for the large number of effect sizes drawn from the same study sample. The results from the multilevel correlated effects analysis yielded a higher average weighted correlation of .13 [.08, .19], $SE = .03$, $p < .0001$. Given the large difference in these estimates, with the results from the multilevel correlated effects analyses yielding an effect size that was nearly twice as large as the model that did not control for study, it appears that not accounting for study dependence drastically impacted our results, which provides support for the methodological decision to also account for study dependence when conducting follow-up moderator analyses.

Looking next at the results of the tests for study heterogeneity, significant heterogeneity was found, with Q [683] = 4639.13, $p < .0001$. The total heterogeneity of the r correlation coefficient was estimated to be high, $I^2 = 85.28\%$. Variance between studies was also found to be significant based on a 95% confidence interval, $\sigma^2 = 0.04$ [0.02-0.06]). Thus, multiple moderator analyses were conducted, one moderator at a time, in order to determine the sample, assessment, and study characteristics that may have significantly contributed to study heterogeneity.

Moderator Analyses for Sample Characteristics

All moderators were entered as categorical, except for age, which was entered as continuous. Age was a significant source of heterogeneity for the correlation between the HME and children's math achievement (F[1, 656] = 10.39, p = .0013, σ^2 = .04 [.02, .06], k = 658) with a -.01 unit decrease in the correlation between the HME and children's math achievement for every 1-year increase in age. The test for residual heterogeneity was significant (Q_E[656] = 2211.72, p < .0001), indicating that even after accounting for age, I^2 = 70.34% of the variability in effect sizes was left unexplained.

All moderation analysis results for sample characteristic moderators are presented in Figure 2. The results of the overall omnibus test with grade as the moderator showed that grade was a significant source of heterogeneity (F[5, 638] = 7.95, p < .0001, σ^2 = .04 [.02, .07], k = 643) in the average weighted correlation between the HME and children's math achievement. Study samples that included only PK/KG children (r = .14 [.08, .20], p < .0001, n = 526) or only elementary school children (r = .12 [.06, .19], p = .0002, n = 75) had average weighted correlations that were positive and significantly greater than zero, but samples that included a combination of PK/KG and elementary school children (r = .20 [-.001, .40], p = .0519, n = 30), a combination of elementary and middle school children (r = .05 [-.21, .30], p = .7204, n = 8), and only middle school children (r = .17 [.23, .57], p = .4044, n = 4) did not. When comparing differences between grades, pairwise t-test results showed that study samples comprised of only elementary school children (b = -0.02 [-0.02, -0.01], p < .0001) demonstrated significantly lower correlations between the HME and children's math achievement than samples made up of only PK/KG children. The test for residual heterogeneity was also significant (Q_E[638] = 4458.82, p <

.0001), indicating that even after accounting for grade, I^2 = 85.69% of the variability in effect sizes was left unexplained.

The results of the overall omnibus test with country as the moderator showed that the study sample's country of origin was a significant source of heterogeneity (F[10, 674] = 6.04, p < .0001, σ^2 = .02 [.01,.04], k = 684) in the correlation between the HME and children's math achievement. Samples from the United States (r = .08 [.02, .14], p = .0083, n = 250), Canada (r = .22 [.09, .34], p = .0005, n = 37), the Netherlands (r = .49 [.33, .66], p < .0001, n = 10), Greece (r = .23 [.09, .38], p = .0014, n = 11), and Australia (r = .26 [0.04, 0.49], p = .0231, n = 6) had average weighted correlations that were positive and significantly different from zero, but samples from Germany (r = .09 [-.22, .39], p = .5792, n = 4), Italy (r = .05 [-.25, .35], p = .7475, n = 21), China (r = .08 [-0.04, 0.19], p = .1809, n = 241), Chile (r = .05 [-0.16, 0.27], p = .6196, n = 36), and the other countries category (r = .03 [-.13, .18], p = .7471, n = 68) did not. When comparing differences between countries of origin, pairwise t-test results showed that, in comparison to the United States, study samples from the Netherlands (b = 0.41 [0.23, 0.59], t(9) = 4.58, p < .0001) demonstrated significantly higher correlations between the HME and children's math achievement, but all other countries of origin did not significantly differ from United States samples, including China (b = -0.004 [-0.12, 0.11], t(9) = -0.06, p = .9508). The test for residual heterogeneity was significant (Q_E[674] = 4348.23, p < .0001), indicating that even after accounting for the sample's country of origin, I^2 = 84.50% of the variability in effect sizes was left unexplained.

Finally, the results of the overall omnibus test with special sample characteristics as the moderator showed that they were a significant source of heterogeneity (F[9, 675] = 3.49, p = .0003, σ^2 = .03 [.02, .06], k = 684) in the average weighted correlation between the HME and children's math achievement. Study samples that were average/typically-developing (r = .16 [.08, .23], p < .0001, n = 135), all one ethnicity (r = .11 [.03, .19], p = .0046, n = 376), all girls (r = .16 [.05, .26], p = .0046, n = 14), all boys (r = .15 [.05, .26], p = .0044, n = 13), or SLI (r = .54 [.17, .91], p = .0044, n = 4) had average weighted correlations that were positive and significantly different from zero, but samples that were low SES (r = .10 [-.01, .22], p = .0725, n = 96), high minority (r = .08 [-0.04, 0.20], p = .1996, n = 18), high SES (r = .13 [-0.25, 0.51], p = .4952, n = 7), or from the other category did not (r = .05 [-.32, .42], p = .7860, n = 7). When comparing across special sample characteristics, pairwise t-tests showed that in comparison to average/typically-developing study samples no significant differences in the average weighted correlation between the HME and children's math achievement were found due to special sample characteristics. Interestingly, low versus high SES samples (b = 0.03 [-0.37, 0.43], t(8) = 0.15, p = .1449), and samples that were made up of all boys versus all girls (b = -0.002 [-0.10, 0.09], t(8) = -0.04, p = .9680) did not show significantly different correlations between the HME and children's math achievement. The test for residual heterogeneity was significant (Q_E[675] = 4489.14, p < .0001), indicating that even after accounting for special sample characteristics, I^2 = 84.96% of the variability in effect sizes was left to be explained.

Moderator Analyses for HME Assessment Characteristics

All moderator results for HME assessment moderators are presented in Figure 3. The results of the overall omnibus test with HME component as the moderator showed that the specific HME component measured was a significant source of heterogeneity (F[8, 676] = 6.47, p < .0001, σ^2 = .01 [.01, .03], k = 684) in the average weighted correlation between the HME and children's math achievement. HME measures that assessed direct HME activities (r = .12 [.06, .18], p < .0001, n = 94), indirect HME activities (r = .06 [.01, .12], p = .0173, n = 199), math expectations (r = .24 [.14, .33], p < .0001, n = 22), parent math talk (r = .14 [.05, .23], p = .0016, n = 210), a combination of direct and indirect HME activities (r = .18 [.10, 026], p < .0001, n = 45), and a combination of activities and attitudes and/or beliefs or expectations (r = .20 [.05, .34], p = .0085, n = 15) had average weighted correlations that were positive and significantly different from zero, but HME measures that assessed parent math attitudes and/or beliefs only (r = .07 [-.001, .14], p = .0540, n = 91) or spatial activities (r = .09 [-.08, .26], p = .2947, n = 8) did not. When comparing between HME components, pairwise t-test results showed that, contrary to our hypothesis, no significant differences in the average weighted correlation between the HME and children's math achievement were found for indirect HME activities (b = 0.01 [-0.01, 0.04], t[7] = 0.85, p = .9517) when compared to direct HME activities. Our pairwise t-test results with parent math expectations as the reference group also violated our hypothesis, showing that none of the other HME components had significantly lower magnitude average weighted correlations with children's math achievement, and instead, HME measures that included a combination of direct and indirect math activities demonstrated significantly higher average weighted correlations (b = 0.08 [0.04, 0.13], t[7] = 3.67, p = .0002) with children's math achievement. The test for residual heterogeneity was not significant (Q_E[676] = 643.71, p = .8090), indicating that after accounting for the specific HME component measured no significant variability in effect sizes remained.

For the HME calculation moderator, the results of the overall omnibus test showed that the specific method of calculation used to measure the HME was a significant source of heterogeneity (F[3, 681] = 11.08, p < .0001, σ^2 = .04 [.02, .06], k = 684) in the average weighted correlation between the HME and children's math achievement. HME calculations that utilized latent factor scores (r = .13 [.07, .19], p < .0001, n = 136), sum scores (r = .16 [.10, .21], p <

.0001, n = 443), or single items (r = .08 [.01, .15], p = .0261, n = 105) had average weighted correlations that were positive and significantly different from zero. When comparing between HME calculation types, with latent factor scores as the reference group, pairwise t-test results showed that no significant differences in the average weighted correlation between the HME and children's math achievement were found (sum scores: b = 0.02 [-0.01, 0.06], t[2] = 1.24, p = .2145; single items: b = -0.06 [-0.11, 0.00], t[2] = -1.83, p = .0675). Thus, contrary to our expectations, HME scores calculated as latent factor scores did not result in significantly higher average weighted correlations between the HME and children's math achievement. The test for residual heterogeneity was significant (Q_E[681] = 2635.78, p < .0001), indicating that even after accounting for the HME calculation method used, I^2 = 85.31% of the variability in effect sizes was left to be explained.

Moderator Analyses for Math Assessment Characteristics

All moderator results for math assessment moderators are presented in Figure 4. The results of the overall omnibus test with math assessment as the moderator showed that the math assessment used to measure children's math achievement was a significant source of heterogeneity (F[13, 671] = 4.06, p < .0001, σ^2 = .04 [.02, .06], k = 684) n the average weighted correlation between the HME and children's math achievement. Researcher-created assessments (r = .13 [.06, .20], p = .0001, n = 416), the KeyMath (r = .10 [.01, .19], p = .0290, n = 13), the WJ (r = .48 [.25, .70], p < .0001, n = 3), the WM (r = .19 [.10, .27], p < .0001, n = 11), the UENT-R (r = .20 [.09, .30], p = .0003, n = 12), the CATM (r = .19 [.07, .32], p = .0030, n = 3), the TEDI-MATH (r = .15 [.04, .26], p = .0075, n = 11), parent-report of children's math achievement (r = .19 [.09, .28], p = .0002, n = 16), or the other math assessment category (r = .12 [.05, .20], p = .0008, n = 95) had average weighted correlations that were positive and significantly different from zero, but the TEMA (r = -.08 [-.25, .08], p = .3174, n = 30), the PENS (r = .16 [-.12, .43], p = .2742, n = 41), or the CMA (r = .29 [-.09, .67], p = .1394, n = 4) did not. Pairwise t-test results showed that, in comparison to researcher-created math assessments, the WJ (b = 0.35 [0.12, 0.57], t[12]= 3.03, p = .0025) had significantly higher average weighted correlations between the HME and children's math achievement, and correlations for the TEMA were significantly lower (b = -0.21 [-0.39, -0.04], t[12] = -2.39, p = .0173). The test for residual heterogeneity was also significant (Q_E[671] = 4480.20, p < .0001), indicating that even after accounting for the specific math assessment used, I^2 = 85.02% of the variability in effect sizes was left to be explained.

The results of the overall omnibus test with math domain as the moderator showed that the math domain assessed was a significant source of heterogeneity (F[5, 679]) = 8.36, p < .0001, σ^2 = .03 [.02, .06], k = 684) in the average weighted correlation between the HME and children's math achievement. When arithmetic operations (r = .12 [.06, .18], p < .0001, n = 117), numerical relations (r = .09 [.03, .15], p = .0024, n = 195), numbering (r = .10 [.04, .16], p = .0014, n = 111), or multiple math domains (r = .14 [.09, .20], p < .0001, n = 257) were assessed the average weighted correlation between the HME and math achievement was positive and significantly different from zero, but when the spatial domain was assessed it was not (r =-.00 [-.14, .14], p = .9919, n = 4). According to pairwise t-test results, in comparison to math measures assessing multiple math domains, all other math domains (numerical relations: b = -0.05 [-0.08, -0.02], p = .0003; numbering: b = -0.05 [-0.08, -0.01], t[4] = -2.67, p = .0077; spatial: b = -0.15 [-0.28, -0.01], t[4] = -2.15, p = .0320), with the exception of arithmetic operations (b = -0.02 [-0.05, 0.01], t[4] = -1.38, p = .1667), had significantly lower average weighted correlations between the HME and children's math achievement. The test for residual heterogeneity was also significant (Q_E[679] = 4597.82, $p < .0001$), indicating that even after accounting for the specific math domain assessed, I^2 = 85.23% of the variability in effect sizes was left unexplained.

For the symbolic math assessment moderator, the results of the overall omnibus test showed that whether math achievement was assessed using a symbolic, non-symbolic, or combination of symbolic and non-symbolic assessments was a significant source of heterogeneity (F [3, 681] = 13.29, $p < .0001$, $\sigma^2 = .04$ [.02, .06], $k = 684$) in the average weighted correlation between the HME and children's math achievement. Symbolic ($r = .15$ [.09, .20], $p < .0001$, $n = 337$), non-symbolic ($r = .12$ [.06, .18], $p < .0001$, $n = 93$), and combined symbolic and non-symbolic ($r = .13$ [.07, .18], $p < .0001$, $n = 254$) math assessments had average weighted correlations that were positive and significantly different from zero. According to pairwise t-tests with symbolic math assessments as the reference group, the average weighted correlation between the HME and children's math achievement was significantly lower when math was assessed using a non-symbolic math assessment ($b = -0.03$ [-0.05, -0.01], $t[3] = -2.41$, $p = .0162$) or a combined symbolic and non-symbolic math assessment ($b = -0.02$ [-0.03, -0.01], $t[3] = -3.68$, $p = .0003$). The test for residual heterogeneity was also significant (Q_E[681] = 4565.45, $p < .0001$), indicating that even after accounting for whether the math assessment was symbolic, I^2 = 85.08% of the variability in effect sizes was left to be explained.

For the timed math assessment moderator, the results of the overall omnibus test showed that whether math achievement was assessed using a timed, untimed, or combination of timed and untimed assessments was a significant source of heterogeneity (F[3, 681] = 14.14, $p < .0001$, $\sigma^2 = .03$ [.02, .06], $k = 684$) in the average weighted correlation between the HME and children's math achievement. Math assessed by measures that were timed ($r = .07$ [.004, .13], $p = .0386$, $n = 67$), untimed ($r = .14$ [.09, .19], $p < .0001$, $n = 607$), or a combination of timed and untimed (r

= .14 [.07, .21], $p < .0001$, $n = 10$) had average weighted correlations that were positive and significantly different from zero. In comparison to math assessed with a timed assessment, pairwise t-tests showed that math assessments that were untimed ($b = 0.07$ [0.04, 0.11], t[3] = 3.88, $p = .0001$) or a combination of timed and untimed ($b = 0.07$ [0.03, 0.12], t[3] = 3.49, $p = .0005$) showed significantly higher average weighted correlations between the HME and children's math achievement. The test for residual heterogeneity was also significant (Q_E[681] = 4601.93, $p < .0001$), indicating that even after accounting for whether the math assessment was timed, $I^2 = 85.20\%$ of the variability in effect sizes was left to be explained.

For the composite math assessment moderator, the results of the overall omnibus test showed that whether math achievement was assessed using a composite or a single math assessment was a significant source of heterogeneity (F[1, 682] = 12.03, $p < .0001$, $\sigma^2 = .04$ [.02, .06], $k = 684$) in the average weighted correlation between the HME and children's math achievement. Both composite ($r = .15$ [.08, .21], $p < .0001$, $n = 71$) and single-measure ($r = .13$ [.08, .18], $p < .0001$, $n = 613$) math assessments had average weighted correlations that were positive and significantly different from zero. Pairwise t-tests, with composite math assessments as the reference group, showed that the average weighted correlation between the HME and children's math achievement was statistically the same when math was assessed as a single measure ($b = -0.02$ [-0.06, 0.02], t[1] = -0.89, $p = .3747$). The test for residual heterogeneity was also significant (Q_E[682] = 4428.74, $p < .0001$), indicating that even after accounting for whether the math assessment was a composite, $I^2 = 84.60\%$ of the variability in effect sizes was left to be explained.

Finally, for the standardized math assessment moderator, the results of the overall omnibus test showed that whether the math assessment was standardized, unstandardized, or a combination of both was a significant source of heterogeneity (F[3, 681] = 9.49, p < .0001, σ^2 = .03 [.02, .06], k = 684) in the average weighted correlation between the HME and children's math achievement. Standardized assessments (r = .15 [.09, .21], p < .0001, n = 85) and unstandardized assessments (r = .12 [.06, .17], p < .0001, n = 585) had average weighted correlations that were positive and significantly different from zero, but combined standardized and unstandardized assessments (r = .31 [-.05, .67], p = .0931, n = 14) did not. Pairwise t-tests with standardized math assessments as the reference group, showed that the average weighted correlation between the HME and children's math achievement were statistically the same for unstandardized (b = -0.03 [-0.06, 0.002], t[2] = -1.82, p = .0695) and combined standardized and unstandardized math assessments (b = 0.16 [-0.1, 0.53], t[2] = 0.86, p = .3874). The test for residual heterogeneity was significant (Q_E[681] = 4386.81, p < .0001), indicating that even after accounting for whether the math assessment was standardized, I^2 = 84.48% of the variability in effect sizes was left to be explained.

Moderator Analyses for Study Characteristics

Moderator results for the study characteristic moderator are presented in Figure 5. The results of the overall omnibus test with whether the study captured longitudinal (at different time points) or concurrent (at the same time point) relations between the HME and children's math achievement as the moderator showed that it was a significant source of heterogeneity (F[2, 682) = 14.50, p < .0001, σ^2 = .04 [.02, .06], k = 684) in the average weighted correlation between the HME and children's math achievement. Both longitudinal studies (r = .08 [.02, .15], p = .0131, n = 109) and studies that measured the HME and children's math achievement concurrently (r = .15 [.10, .21], p < .0001) had average weighted correlations that were positive and significantly different from zero. Pairwise t-tests with longitudinal studies as the reference group showed that effect sizes that captured concurrent relations had average weighted correlations between the HME and children's math achievement that were significantly higher (b = 0.07 [0.02, 0.12], p = .0124, n = 575). The test for residual heterogeneity was also significant (Q_E[682] = 4635.70, p < .0001), indicating that even after accounting for whether the math assessment was standardized, I^2 = 85.29% of the variability in effect sizes was left to be explained.

Overall, results from all 14 individual omnibus tests, except for the test including the HME component moderator, showed significant residual heterogeneity remained after accounting for the moderator modeled. This means that variability in the observed effect sizes was significantly larger than would be induced by sampling error alone, and that other moderators not tested in each moderator model were influencing the magnitude of the correlation between the HME and children's math achievement. Given that each moderator was tested individually, it is not surprising that no single moderator (with the exception of the HME component moderator) accounted for all significant effect size variance. Thus, as a final step to determine the amount of study heterogeneity accounted for by all potential moderators at once, a multilevel correlated effects meta-analysis was run with all 14 coded sample, assessment, and study features included. Results from the overall omnibus test that included all moderators showed that the combined moderators were a significant source of heterogeneity (F[55, 561] = 4.04, p < .0001, σ^2 = .04 [.01, .09] in the average weighted correlation between the HME and children's math achievement. The residual heterogeneity of the model including all moderators was also significant (Q_E(561) = 1358.52, p < .0001, k = 617), with I^2 = 58.71% of the variability in effect sizes left unexplained, indicating that other moderators not tested in the present analysis were likely influencing the magnitude of the correlation between the HME and children's math achievement, beyond the effects of sampling error.

Publication Bias

Funnel plot. First, publication bias was assessed using a funnel plot of effect sizes to standard errors, which is depicted in Figure 6. A visual inspection showed multiple studies outside of the shaded area, suggesting slight publication bias in the negative direction, indicating that more studies with larger p-values are likely missing from the literature on the HME and children's math achievement. The results from Egger's test, which provides a parametric test for the skew of the distribution of effect sizes, confirmed the existence of significant publication bias ($z = -2.71$, $p = .0067$), with slightly more effect sizes falling below the mean than above it. Given that slightly more lower magnitude correlations are reported than are higher magnitude correlations (i.e., above the meta-analytic average), our results do not support the existence of a file-drawer problem, wherein small effect sizes fail to be published and/or reported.

P-curve analysis. Results from the continuous p-curve analysis showed that both the full ($Z = -5.27$, $p < .0001$) and half ($Z = -5.02$, $p < .0001$) p-curve tests supported the existence of a significant right skew (see Figure 7). These combination test results, which have been shown to be more robust to p-hacking than a simple p-curve test (Simonsohn et al., 2014), indicated that the set of significant findings had evidential value. Furthermore, full p-curve, and both the half p-curve and binomial 33% power test were non-significant (full: $Z = 2.90$, $p > .9999$; half: $Z = 4.24$, $p > .9999$; binomial: $p > .9999$), indicating that the p-curve does not support that the evidential value is inadequate nor absent. These combined results indicate that the present meta-analytic sample of studies has evidential value and does not show evidence of *p*-hacking.

Sensitivity Analyses

Trim-and-fill. Trim-and-fill results are presented in Figure 8. The trim-and-fill procedure estimated that four studies were missing from above the 99% confidence interval around the average weighted correlation between the HME and children's math achievement. The empty circles represent the filled-in studies ($n = 4$), and the black circles represent the studies that were not trimmed from the analysis (in this case, no studies were trimmed from analysis). A meta-analysis on these included hypothesized studies resulted in the same estimate of r in comparison to the random-effects analysis results that did not utilize multilevel correlated effects modeling ($r = 0.07$ [0.05, 0.09], $p < .0001$).

Fail-Safe N. According to the results of the fail-safe N test using the Rosenthal approach, in order to achieve null population results (i.e., $r = 0$), an additional 104,948 studies with null results, showing no significant association between the HME and children's math achievement, are needed to achieve the target null p-value of $> .05$. To achieve a $p > .01$, an additional 52,124 studies with null results ($r = 0$) are needed. Given that so many studies would have to be added in order to support the null hypothesis that no significant association exists between the HME and children's math achievement, the fail-safe N test results provide evidence that our meta-analytic results are robust and that our study sample has evidential value.

Robust Variance Estimation. The results of the RVE analyses using robumeta indicated that the effect sizes, standard errors, and τ^2 values were robust across different values of ρ ($r = .14$, $SE = .03$, $\tau^2 = .03$ for all values of ρ). Although the estimate of the average weighted correlation was slightly higher than our meta-analytic results using the metafor package of $r = .13$, the fact that the results are almost the same, with overlapping confidence intervals, indicates that our meta-analytic results using metafor are also robust.

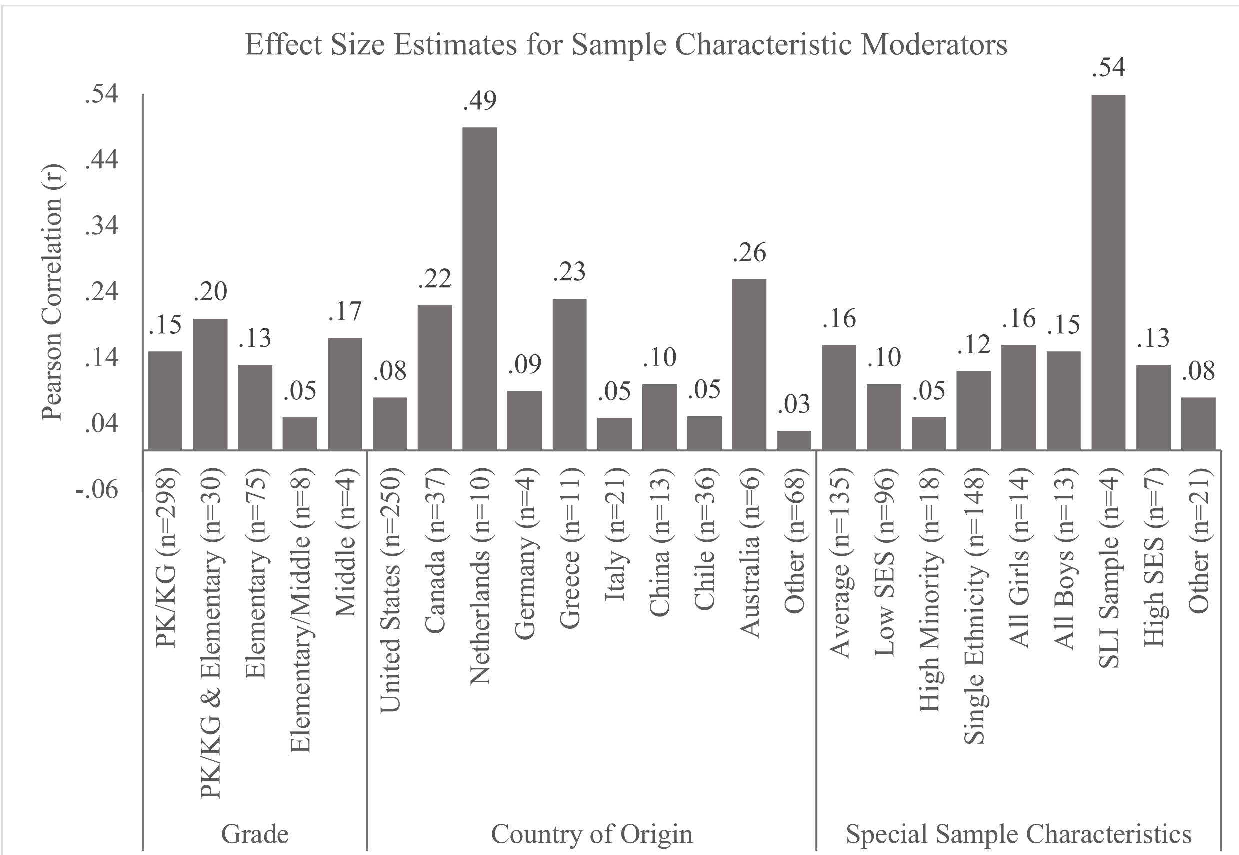

Figure 9. Effect size estimates for sample characteristic moderators, including grade, country of origin, and special population characteristics.

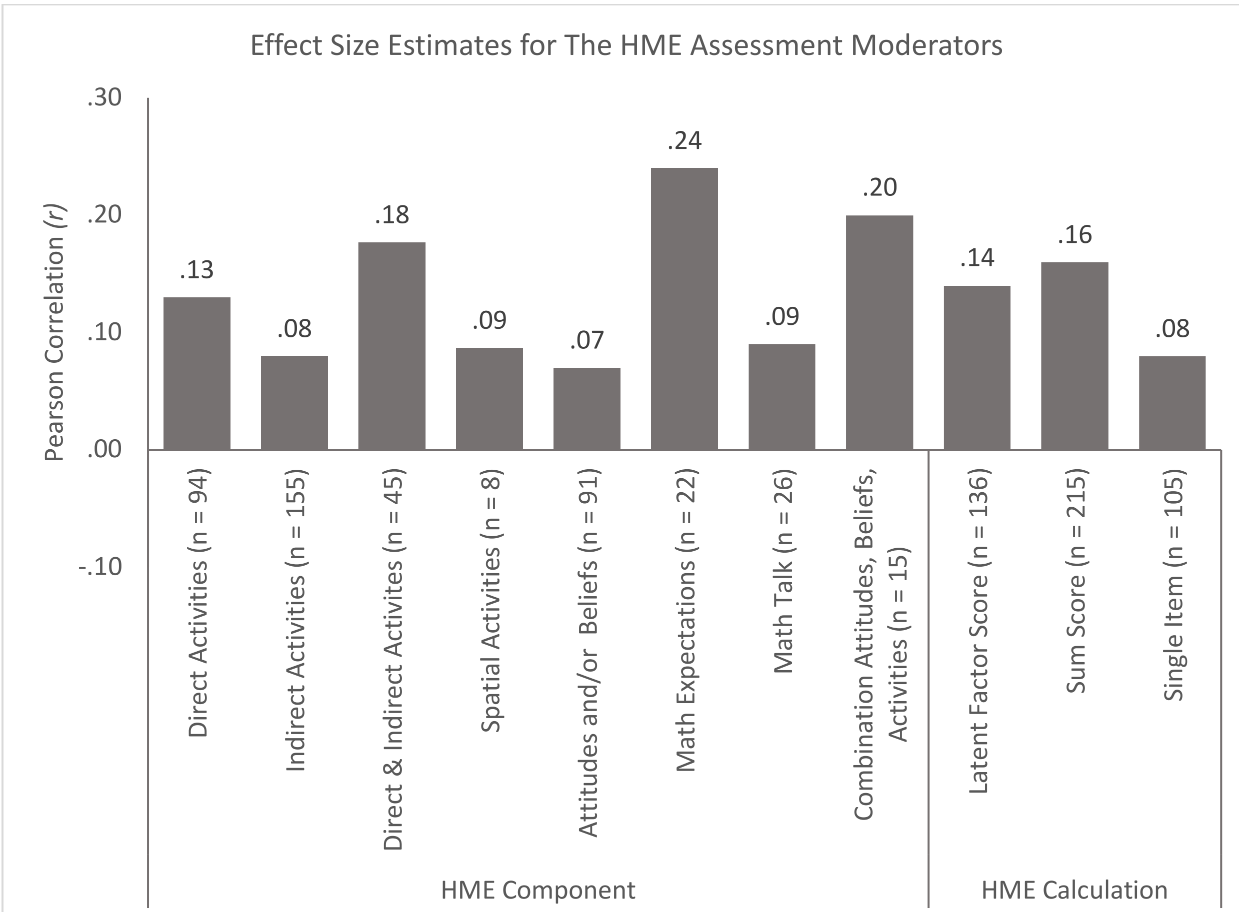

Figure 10. Effect size estimates for HME assessment moderators, including HME component and HME calculation.

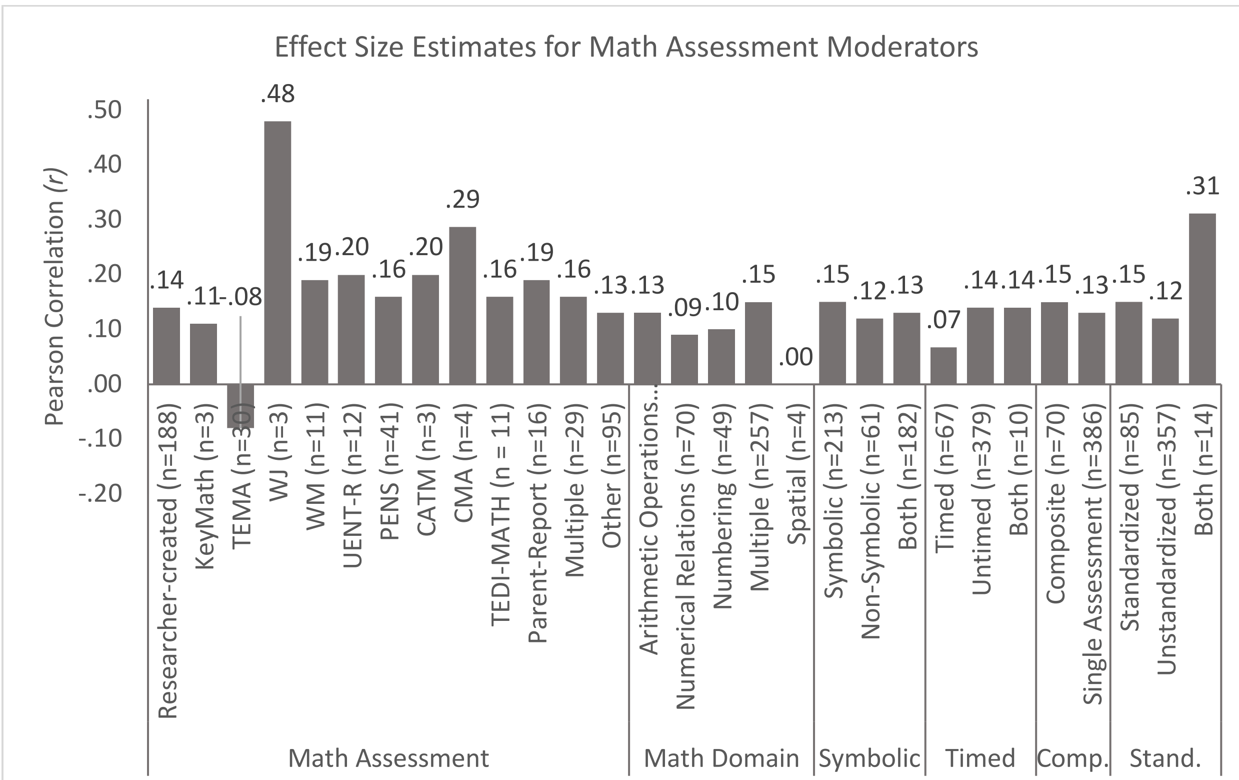

Figure 11. Effect size estimates for math assessment moderators, including math assessment, math domain, symbolic or non-symbolic, timed or untimed, composite or single assessment (Comp.), and standardized or unstandardized (Stand.).

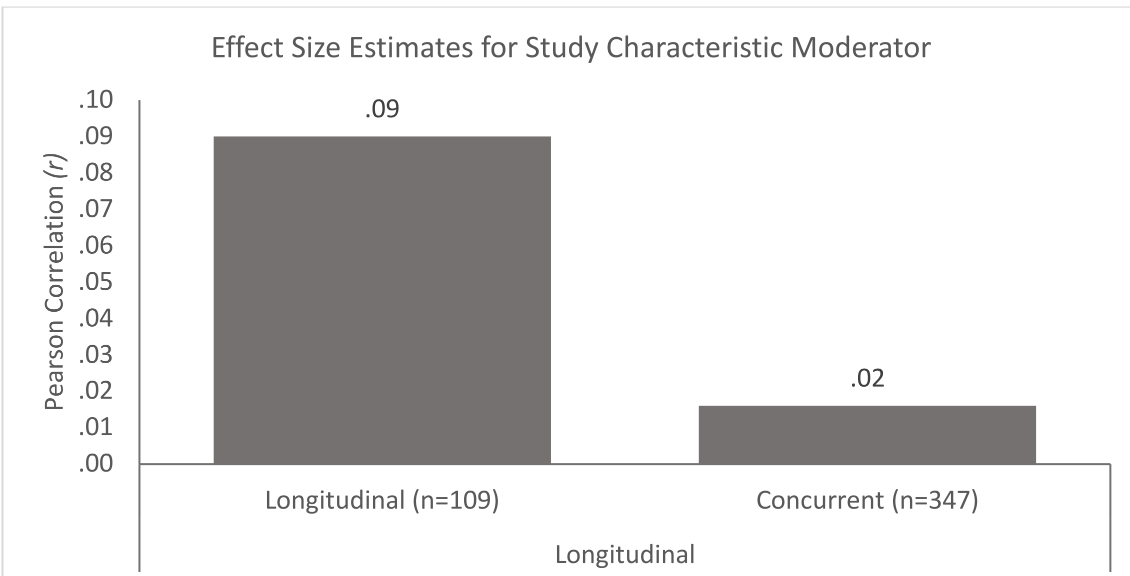

Figure 12. Effect size estimates for the study characteristic moderator of whether the study reported longitudinal or concurrent effect sizes.

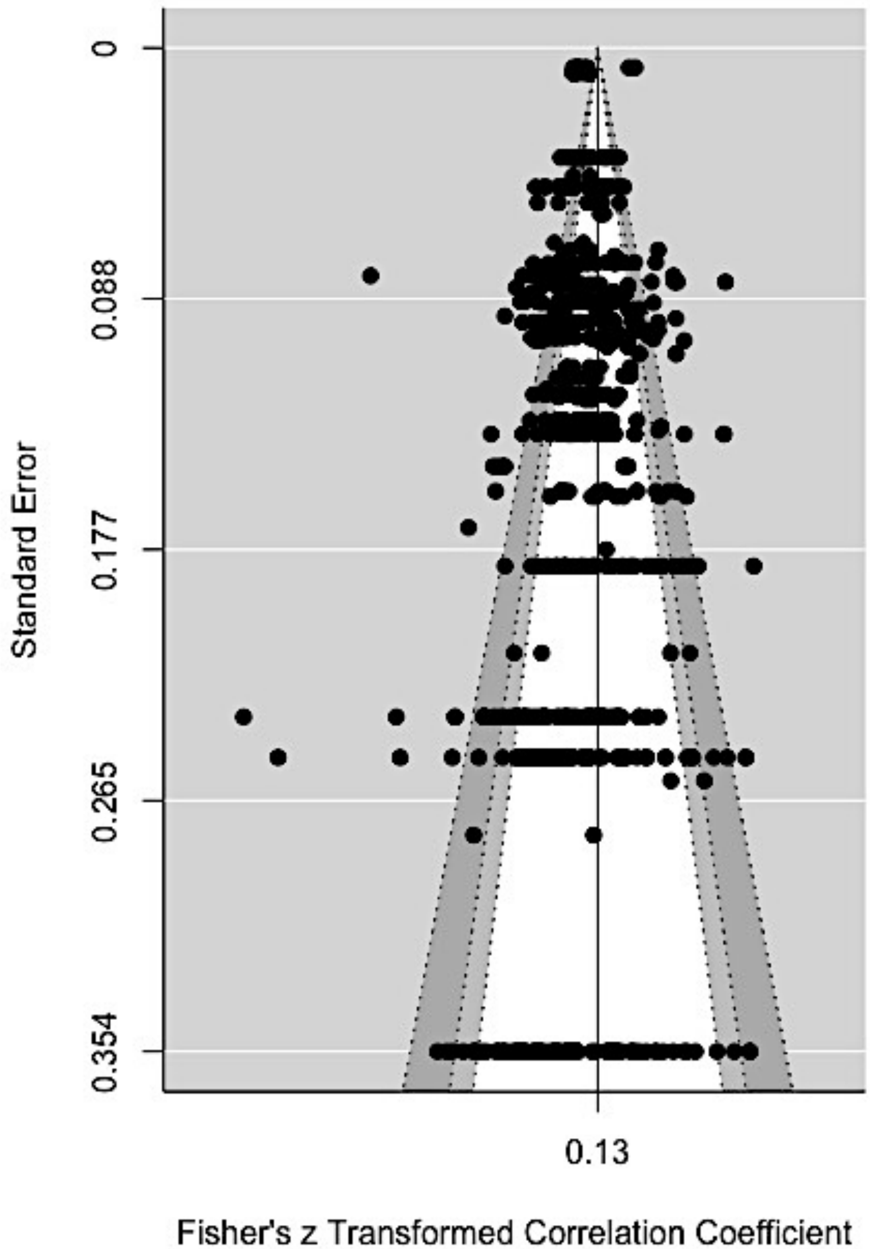

Figure 13. Enhanced funnel plot of the multilevel correlated effects meta-analysis results with the average weighted correlation of $r = .13$ and confidence intervals on the 90^{th} (white), 95^{th} (gray), and 99^{th} (dark gray) percentiles.

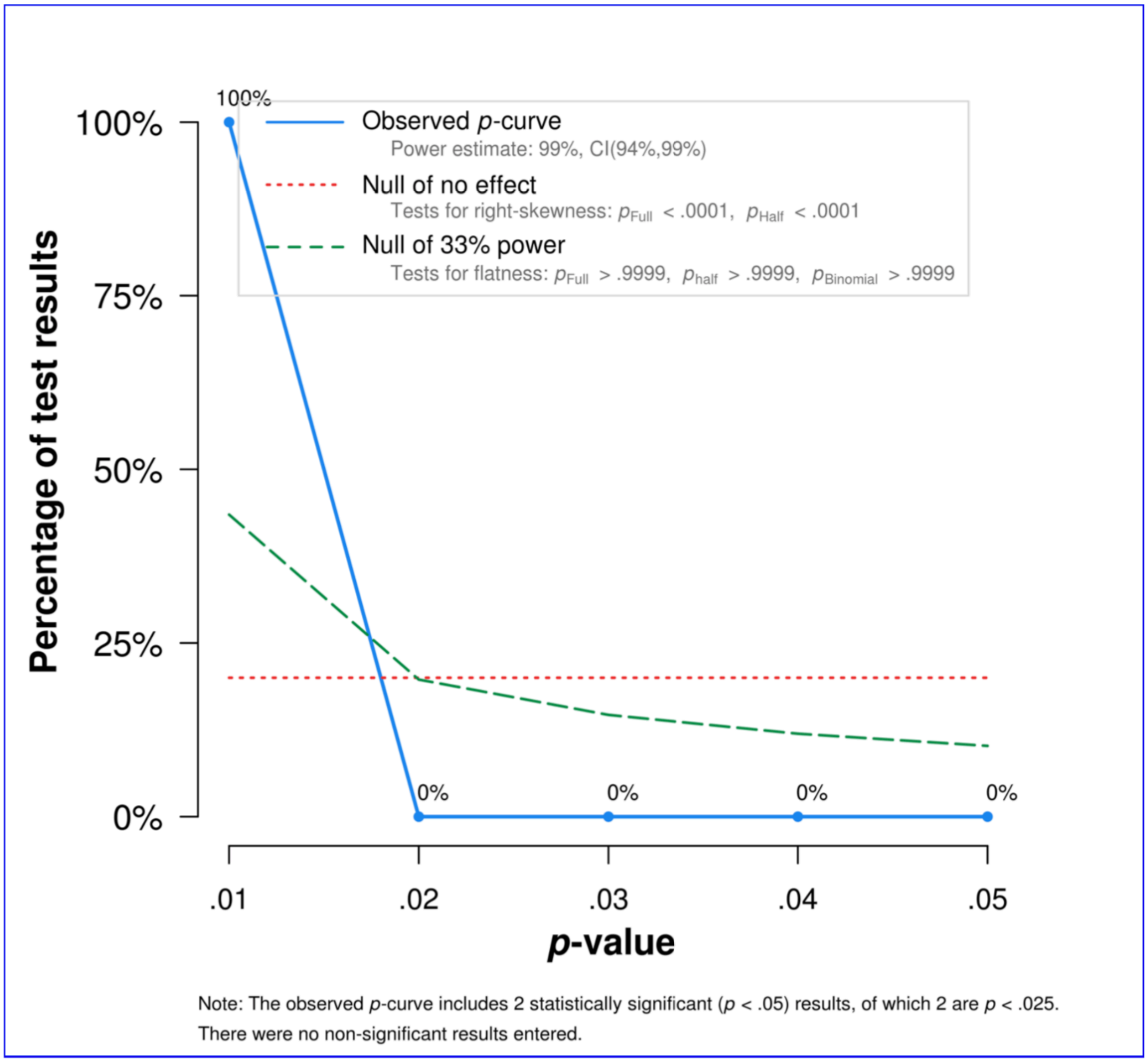

Figure 14. P-curve analysis results showing that a significant right skew and non-significant results at 33% power for both the full and half p-curve tests, indicating evidential value and a lack of *p*-hacking in our study sample.

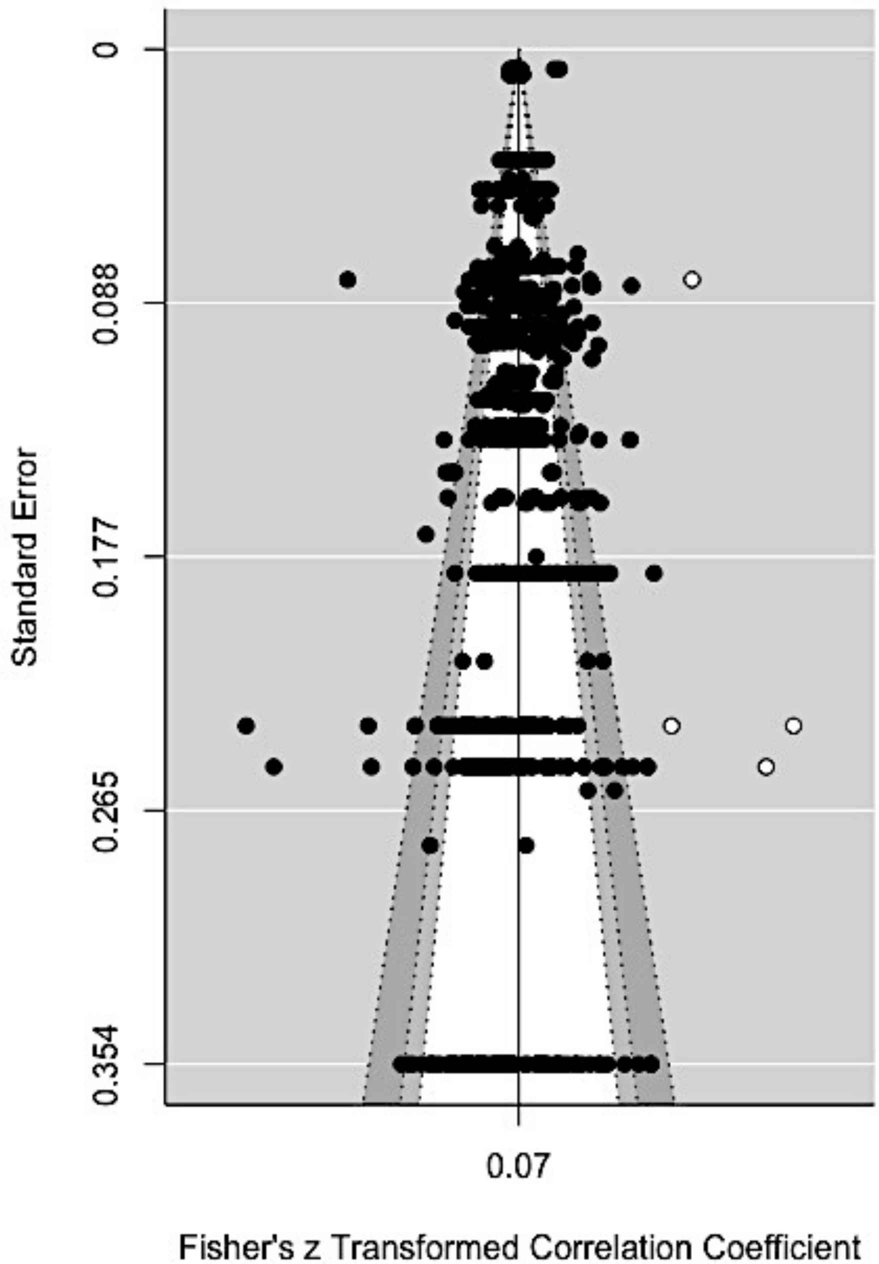

Figure 15. Trim-and-fill plot of the random effects meta-analysis average weighted correlation of $r = .07$ and confidence intervals on the 90^{th} (white), 95^{th} (gray), and 99^{th} (dark gray) percentiles. The trim-and-fill method does not allow for multilevel analyses, so the analysis was conducted without accounting for effect size study dependence ($r = .07$).

REFERENCES

* Indicates references included in the meta-analytic sample

Aunola, K., Leskinen, E., Lerkkanen, M. K., & Nurmi, J. E. (2004). Developmental dynamics of math performance from preschool to grade 2. *Journal of Educational Psychology*, *96*(4), 699.

Aunio, P., & Niemivirta, M. (2010). Predicting children's mathematical performance in grade one by early numeracy. *Learning and Individual Differences*, *20*, 427– 435.

*Benavides-Varela, S., Butterworth, B., Burgio, F., Arcara, G., Lucangeli, D., & Semenza, C. (2016). Numerical activities and information learned at home link to the exact numeracy skills in 5–6 years-old children. *Frontiers in Psychology*, *7*, 94.

*Benigno, J. P., & Ellis, S. (2004). Two is greater than three: Effects of older siblings on parental support of preschoolers' counting in middle-income families. *Early Childhood Research Quarterly*, *19*(1), 4-20.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289-300.

*Bhanot, R., & Jovanovic, J. (2005). Do parents' academic gender stereotypes influence whether they intrude on their children's homework?. *Sex Roles*, *52*(9-10), 597-607.

*Blevins-Knabe, B., & Musun-Miller, L. (1996). Number use at home by children and their parents and its relationship to early mathematical performance. *Early Development and Parenting: An International Journal of Research and Practice*, *5*(1), 35-45.

Bornstein, M. H., & Cheah, C. S. (2006). The place of "culture and parenting" in the ecological contextual perspective on developmental science. *Parenting Beliefs, Behaviors, and Parent-Child Relations: A Cross-Cultural Perspective*, 3-33.

Borenstein, M., Cooper, H., Hedges, L., & Valentine, J. (2009). Effect sizes for continuous data. *The handbook of research synthesis and meta-analysis*, *2*, 221-235.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to meta-analysis*. John Wiley & Sons.

Bracken, B. A. (2006). Bracken Basic Concepts Scale–Revised. San Antonio, TX: Psychological Corporation.

Bronfenbrenner, U. (1979). Contexts of child rearing: Problems and prospects. *American Psychologist*, *34*(10), 844.

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

Burgess, S. R., Hecht, S. A., & Lonigan, C. J. (2002). Relations of the home literacy environment (HLE) to the development of reading-related abilities: A one-year longitudinal study. *Reading Research Quarterly*, *37*(4), 408-426.

Bus, A. G., Van Ijzendoorn, M. H., & Pellegrini, A. D. (1995). Joint book reading makes for success in learning to read: A meta-analysis on intergenerational transmission of literacy. *Review of Educational Research*, *65*(1), 1-21.

*Cai, J. (2003). Investigating parental roles in students' learning of mathematics from a cross-national perspective. *Mathematics Education Research Journal*, *15*(2), 87-106.

*Cain-Caston, M. (1993). Parent and student attitudes toward mathematics as they relate to third grade mathematics achievement. *Journal of Instructional Psychology*, *20*(2), 96.

Card, N. A. (2011). Applied meta-analysis for social science research. New York: Guilford Press.

Casey, B. M., Dearing, E., Dulaney, A., Heyman, M., & Springer, R. (2014). Young girls' spatial and arithmetic performance: The mediating role of maternal supportive interactions during joint spatial problem solving. *Early Childhood Research Quarterly*, *29*(4), 636-648.

*Cheung, S. K. (2012). *Promoting Preschoolers' Number Sense and Interest in Learning Mathematics: How Should Parents Interact with Their Children During Home Numeracy Activities?* (Doctoral dissertation, Chinese University of Hong Kong).

*Cheung, S. K., Yang, X., Dulay, K. M., & McBride, C. (2018). Family and individual variables associated with young Filipino children's numeracy interest and competence. *British Journal of Developmental Psychology*, *36*(2), 334-353.

Cicchetti, D. V., & Sparrow, S. A. (1981). Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. *American Journal of Mental Deficiency*, *86*(2), 127-137.

*Ciping, D., Silinskas, G., Wei, W., & Georgiou, G. K. (2015). Cross-lagged relationships between home learning environment and academic achievement in Chinese. *Early Childhood Research Quarterly*, *33*, 12-20.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37-46.

Cohrssen, C., Church, A., & Tayler, C. (2014). Pausing for learning: Responsive engagement in mathematics activities in early childhood settings. *Australasian Journal of Early Childhood*, *39*(4), 95-102.

Connolly, A. J. (2007). *KeyMath 3: Diagnostic Assessment*. San Antonio, TX: Pearson.

Crosnoe, R., Leventhal, T., Wirth, R. J., Pierce, K. M., Pianta, R. C., & NICHD Early Child Care Research Network. (2010). Family socioeconomic status and consistent environmental stimulation in early childhood. *Child Development*, *81*(3), 972-987.

*Dearing, E., Casey, B. M., Ganley, C. M., Tillinger, M., Laski, E., & Montecillo, C. (2012). Young girls' arithmetic and spatial skills: The distal and proximal roles of family socioeconomics and home learning experiences. *Early Childhood Research Quarterly*, *27*(3), 458-470.

*DeFlorio, L., & Beliakoff, A. (2015). Socioeconomic status and preschoolers' mathematical knowledge: The contribution of home activities and parent beliefs. *Early Education and Development*, *26*(3), 319-341.

*Del Río, M. F., Susperreguy, M. I., Strasser, K., & Salinas, V. (2017). Distinct influences of mothers and fathers on kindergartners' numeracy performance: The role of math anxiety, home numeracy practices, and numeracy expectations. *Early Education and Development*, *28*(8), 939-955.

*Drummond, K. V., & Stipek, D. (2004). Low-income parents' beliefs about their role in children's academic learning. *The Elementary School Journal*, *104*(3), 197-213.

Duncan, G. J., & Magnuson, K. A. (2005). Can family socioeconomic resources account for racial and ethnic test score gaps?. *The Future of Children*, 35-54.

Duncan, G. J., & Magnuson, K. (2011). The nature and impact of early achievement skills, attention skills, and behavior problems. *Whither Opportunity*, *15*(1), 47-70.

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

Duval, S., & Tweedie, R. (2000a). Trim and fill: a simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455-463.

Duval, S., & Tweedie, R. (2000b). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, *95*(449), 89-98.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *Bmj*, *315*(7109), 629-634.

*Elliott, L., Braham, E. J., & Libertus, M. E. (2017). Understanding sources of individual variability in parents' number talk with young children. *Journal of Experimental Child Psychology*, *159*, 1-15.

*Else-Quest, N. M., Hyde, J. S., & Hejmadi, A. (2008). Mother and child emotions during mathematics homework. *Mathematical Thinking and Learning*, *10*(1), 5-35.

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: a meta-analysis. *Psychological Bulletin*, *136*(1), 103.

Fleiss, J. L. (1981). Balanced incomplete block designs for inter-rater reliability studies. *Applied Psychological Measurement*, *5*(1), 105-112.

Foster, T. D., Froyen, L. C., Skibbe, L. E., Bowles, R. P., & Decker, K. B. (2016). Fathers' and mothers' home learning environments and children's early academic outcomes. *Reading and Writing*, *29*(9), 1845-1863.

Galindo, C., & Sheldon, S. B. (2012). School and home connections and children's kindergarten achievement gains: The mediating role of family involvement. *Early Childhood Research Quarterly*, *27*(1), 90-103.

Gayán, J., & Olson, R. K. (2003). Genetic and environmental influences on individual differences in printed word recognition. *Journal of Experimental Child Psychology*, *84*(2), 97-123.

Geary, D. C. (2004). Mathematics and learning disabilities. *Journal of Learning Disabilities*, *37*(1), 4-15.

Ginsburg, H., & Baroody, A. J. (2003). *TEMA-3: Test of early mathematics ability*. Pro-ed.

Greenwald, R., Hedges, L. V., & Laine, R. D. (1996). The effect of school resources on student achievement. *Review of Educational Research*, *66*(3), 361-396.

Grégoire, J., Noël, M., & Van Nieuwenhoven, C. (2004). Tedi-Math. *Antwerpen: Harcourt*.

Hart, S. A., Petrill, S. A., Thompson, L. A., & Plomin, R. (2009). The ABCs of math: A genetic analysis of mathematics and its links with reading ability and general cognitive ability. *Journal of Educational Psychology*, *101*(2), 388.

*Hart, S. A., Ganley, C. M., & Purpura, D. J. (2016). Understanding the home math environment and its role in predicting parent report of children's math skills. *PloS one*, *11*(12), e0168227.

Head, M. L., Holman, L., Lanfear, R., Kahn, A. T., & Jennions, M. D. (2015). The extent and consequences of p-hacking in science. *PLoS Biology*, *13*(3), e1002106.

Hedges, L. V. (1983). A random effects model for effect sizes. *Psychological Bulletin*, *93*(2), 388.

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, *1*(1), 39-65.

Hindman, A. H., Skibbe, L. E., Miller, A., & Zimmerman, M. (2010). Ecological contexts and early learning: Contributions of child, family, and classroom factors during Head Start, to literacy and mathematics growth through first grade. *Early Childhood Research Quarterly*, *25*(2), 235-250.

Hojnoski, R. L., & Floyd, R. G. (2013). Individual growth and development indicators of early numeracy (IGDIs-EN). Minneapolis, MN: Early Learning Labs.

*Holod, A. (2012). *Parental Time or Money: What Matters More for Children's School Success?* (Doctoral dissertation, Teachers College).

Huntsinger, C. S., Jose, P. E., Larson, S. L., Balsink Krieg, D., & Shaligram, C. (2000). Mathematics, vocabulary, and reading development in Chinese American and European American children over the primary school years. *Journal of Educational Psychology*, *92*(4), 745.

*Huntsinger, C. S., Jose, P. E., & Luo, Z. (2016). Parental facilitation of early mathematics and reading skills and knowledge through encouragement of home-based activities. *Early Childhood Research Quarterly*, *37*, 1-15.

Ikäheimo, H. (1996). Diagnostic test for basic mathematical concepts. Helsinki, Finland: Oy OPPERI Ab.

Jacobs, N., & Harvey, D. (2005). Do parents make a difference to children's academic achievement? Differences between parents of higher and lower achieving students, *Educational Studies*, *31*(4), 431-448.

Jordan, N. C., Kaplan, D., Locuniak, M. N., & Ramineni, C. (2007). Predicting first-grade math achievement from developmental number sense trajectories. *Learning Disabilities Research & Practice*, *22*(1), 36-46.

Keith, P. B., & Lichtman, M. V. (1994). Does parental involvement influence the academic achievement of Mexican-American eighth graders? Results from the National Education Longitudinal Study. *School Psychology Quarterly*, *9*(4), 256.

Kellaghan, T., Sloane, K., Alvarez, B., & Bloom, B. S. (1993). *The home environment and school learning: Promoting parental involvement in the education of children*. Jossey-Bass.

*Kleemans, T., Peeters, M., Segers, E., & Verhoeven, L. (2012). Child and home predictors of early numeracy skills in kindergarten. *Early Childhood Research Quarterly*, *27*(3), 471-477.

*Kleemans, T., Segers, E., & Verhoeven, L. (2013). Relations between home numeracy experiences and basic calculation skills of children with and without specific language impairment. *Early Childhood Research Quarterly*, *28*(2), 415-423.

Klein, A., Starkey, P., Clements, D., Sarama, J., & Iyer, R. (2008). Effects of a pre-kindergarten mathematics intervention: A randomized experiment. *Journal of Research on Educational Effectiveness*, *1*(3), 155-178.

Klein, A., Starkey, P., & Wakeley, A. (2000). Child math assessment: Preschool battery (cma). *Berkeley, CA: University of California*.

Knapp, G., & Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics In Medicine*, *22*(17), 2693-2710.

Konstantopoulos, S. (2011). Fixed effects and variance components estimation in three-level meta-analysis. *Research Synthesis Methods*, *2*(1), 61-76.

Krajewski, K., Küspert, P., Schneider, W., & Visé, M. (2002). DEMAT 1+. *Deutscher Mathematiktest für erste Klassen. Göttingen: Hogrefe*.

*LeFevre, J. A., Skwarchuk, S. L., Smith-Chant, B. L., Fast, L., Kamawar, D., & Bisanz, J. (2009). Home numeracy experiences and children's math performance in the early school years. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, *41*(2), 55.

*LeFevre, J. A., Polyzoi, E., Skwarchuk, S. L., Fast, L., & Sowinski, C. (2010). Do home numeracy and literacy practices of Greek and Canadian parents predict the numeracy skills of kindergarten children?. *International Journal of Early Years Education*, *18*(1), 55-70.

*Levine, S. C., Suriyakham, L. W., Rowe, M. L., Huttenlocher, J., & Gunderson, E. A. (2010). What counts in the development of young children's number knowledge?. *Developmental Psychology*, *46*(5), 1309.

Lichtenberger, E. O. (2008). Stanford Diagnostic Mathematics Test–Fourth Edition (SDMT4). *Encyclopedia of Special Education*, 1915-1915.

*Lindberg, S. M., Hyde, J. S., & Hirsch, L. M. (2008). Gender and mother-child interactions during mathematics homework: The importance of individual differences. *Merrill-Palmer Quarterly* *54*(2), 232-255.

*Manolitsis, G., Georgiou, G. K., & Tziraki, N. (2013). Examining the effects of home literacy and numeracy environment on early reading and math acquisition. *Early Childhood Research Quarterly*, *28*(4), 692-703.

Mboya, M. M. (1986). Black adolescents: A descriptive study of their self-concepts and academic achievement. *Adolescence*, *21*(83), 689.

Methe, S. A., Hintze, J. M., & Floyd, R. G. (2008). Validation and decision accuracy of early numeracy skill indicators. *School Psychology Review*, *37*(3), 359.

*Missall, K., Hojnoski, R. L., Caskie, G. I., & Repasky, P. (2015). Home numeracy environments of preschoolers: Examining relations among mathematical activities, parent mathematical beliefs, and early mathematical skills. *Early Education and Development*, *26*(3), 356-376.

Muñoz-Sandoval, A. F., Woodcock, R. W., McGrew, K. S., & Mather, N. (2005). Batería III Pruebas de aprovechamiento. *Itasca, IL: Riverside Publishing*.

*Mutaf Yıldız, B., Sasanguie, D., De Smedt, B., & Reynvoet, B. (2018). Frequency of home numeracy activities is differentially related to basic number processing and calculation skills in kindergartners. *Frontiers in Psychology*, *9*, 340.

*Mutaf Yildiz, B., Sasanguie, D., De Smedt, B., & Reynvoet, B. (2018). Investigating the relationship between two home numeracy measures: a questionnaire and observations during Lego building and book reading. *British Journal of Developmental Psychology*, *36*(2), 354-370.

*Napoli, A. R., & Purpura, D. J. (2018). The home literacy and numeracy environment in preschool: Cross-domain relations of parent–child practices and child outcomes. *Journal of Experimental Child Psychology*, *166*, 581-603.

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel.* Retrieved on July 16, 2008, from http://www.ed.gov/about/bdscomm/list/ mathpanel/report/final-report.pdf

National Research Council. (2009). Mathematics learning in early childhood: Paths toward excellence and equity. Washington, DC: Author.

*Niklas, F., & Schneider, W. (2014). Casting the die before the die is cast: The importance of the home numeracy environment for preschool children. *European Journal of Psychology of Education*, *29*(3), 327-345.

*Niklas, F., Cohrssen, C., & Tayler, C. (2016). Improving preschoolers' numerical abilities by enhancing the home numeracy environment. *Early Education and Development*, *27*(3), 372-383.

*Niklas, F., Cohrssen, C., & Tayler, C. (2016). Parents supporting learning: a non-intensive intervention supporting literacy and numeracy in the home learning environment. *International Journal of Early Years Education*, *24*(2), 121-142.

Passolunghi, M. C., Lanfranchi, S., Altoè, G., & Sollazzo, N. (2015). Early numerical abilities and cognitive skills in kindergarten children. *Journal of Experimental Child Psychology*, *135*, 25-42.

Petrill, S., Logan, J., Hart, S., Vincent, P., Thompson, L., Kovas, Y., & Plomin, R. (2012). Math fluency is etiologically distinct from untimed math performance, decoding fluency, and untimed reading performance: Evidence from a twin study. *Journal of Learning Disabilities*, *45*(4), 371-381.

*Pezdek, K., Berry, T., & Renno, P. A. (2002). Children's mathematics achievement: The role of parents' perceptions and their involvement in homework. *Journal of Educational Psychology*, *94*(4), 771.

*Phillipson, S., & Phillipson, S. N. (2007). Academic expectations, belief of ability, and involvement by parents as predictors of child achievement: A cross-cultural comparison. *Educational Psychology*, *27*(3), 329-348.

Prokic-Breuer, T., & Dronkers, J. (2012). The high performance of Dutch and Flemish 15-year-old native pupils: Explaining country differences in math scores between highly stratified educational systems. *Educational Research and Evaluation*, *18*(8), 749-777.

Provasnik, S., Kastberg, D., Ferraro, D., Lemanski, N., Roey, S., and Jenkins, F. (2012). *Highlights From TIMSS 2011: Mathematics and Science Achievement of U.S. Fourth- and Eighth-Grade Students in an International Context* (NCES 2013- 009). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

*Puccioni, J. (2015). Parents' conceptions of school readiness, transition practices, and children's academic achievement trajectories. *The Journal of Educational Research*, *108*(2), 130-147.

Purpura, D. J. (2009). Informal number-related mathematics skills: An examination of the structure of and relations between these skills in preschool. *Unpublished dissertation, Florida State University*.

Purpura, D. J., & Lonigan, C. J. (2013). Informal numeracy skills: The structure and relations among numbering, relations, and arithmetic operations in preschool. *American Educational Research Journal*, *50*(1), 178-209.

Purpura, D. J., Reid, E. E., Eiland, M. D., & Baroody, A. J. (2015). Using a brief preschool early numeracy skills screener to identify young children with mathematics difficulties. *School Psychology Review*, *44*(1), 41-59.

Purpura, D. J., Schmitt, S. A., & Ganley, C. M. (2017). Foundations of mathematics and literacy: The role of executive functioning components. *Journal of Experimental Child Psychology*, *153*, 15-34.

*Ramani, G. B., Rowe, M. L., Eason, S. H., & Leech, K. A. (2015). Math talk during informal learning activities in Head Start families. *Cognitive Development*, *35*, 15-33.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638.

Sarama, J., & Clements, D. H. (2009). Early childhood mathematics education research: learning trajectories for young children. New York, NY: Routledge.

*Segers, E., Kleemans, T., & Verhoeven, L. (2015). Role of parent literacy and numeracy expectations and activities in predicting early numeracy skills. *Mathematical Thinking and Learning*, *17*(2-3), 219-236.

Sénéchal, M., & LeFevre, J. A. (2002). Parental involvement in the development of children's reading skill: A five-year longitudinal study. *Child Development*, *73*(2), 445-460.

*Silinskas, G., Leppänen, U., Aunola, K., Parrila, R., & Nurmi, J. E. (2010). Predictors of mothers' and fathers' teaching of reading and mathematics during kindergarten and Grade 1. *Learning and Instruction*, *20*(1), 61-71.

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2015). Specification curve: Descriptive and inferential statistics on all reasonable specifications. doi: 10.2139/ssrn.2694998

*Skwarchuk, S. L. (2009). How do parents support preschoolers' numeracy learning experiences at home?. *Early Childhood Education Journal*, *37*(3), 189-197.

*Skwarchuk, S. L., Sowinski, C., & LeFevre, J. A. (2014). Formal and informal home learning activities in relation to children's early numeracy and literacy skills: The development of a home numeracy model. *Journal of Experimental Child Psychology*, *121*, 63-84.

*Susperreguy, M. I., & Davis-Kean, P. E. (2016). Maternal math talk in the home and math skills in preschool children. *Early Education and Development*, *27*(6), 841-857.

*Susperreguy, M. I., Douglas, H., Xu, C., Molina-Rojas, N., & LeFevre, J. A. (2018). Expanding the Home Numeracy Model to Chilean children: Relations among parental expectations, attitudes, activities, and children's mathematical outcomes. *Early Childhood Research Quarterly*.

*Thompson, R. J., Napoli, A. R., & Purpura, D. J. (2017). Age-related differences in the relation between the home numeracy environment and numeracy skills. *Infant and Child Development*, *26*(5), e2019.

Tiegs, E., & Clark, W. (1977). California achievement test. Monterey, Calif.: CTB.
Tymms, P., & Albone, S. (2002). Performance indicators in primary schools. *School improvement through performance feedback*, 191-218.

Tymms, P. and Merrell, C. and Hawker, D., & Nicholson, F. (2014). “Performance indicators in primary schools: A comparison of performance on entry to school and the progress made in the first year in England and four other jurisdictions”. Research report, Project Report. Department for Education, London.

van Luit, H., & van de Rijt, B. (2009). *Utrechtse getalbegrip toets-revised*. Graviant Educatieve Uitgaven.

Van Luit, J. E. H., Van de Rijt, B. A. M., & Pennings, A. H. (1994). Utrechtse getalbegrip toets [early numeracy test]. *Doetinchem, The Netherlands: Graviant*.

*Vandermaas-Peeler, M., & Pittard, C. (2014). Influences of social context on parent guidance and low-income preschoolers' independent and guided math performance. *Early Child Development and Care*, *184*(4), 500-521.

*Vandermaas-Peeler, M., Boomgarden, E., Finn, L., & Pittard, C. (2012). Parental support of numeracy during a cooking activity with four-year-olds. *International Journal of Early Years Education*, *20*(1), 78-93.

Viechtbauer, W. (2010). Metafor: meta-analysis package for R. *R package version*, *2010*, 1-0.
*Vukovic, R. K., Roberts, S. O., & Green Wright, L. (2013). From parental involvement to children's mathematical performance: The role of mathematics anxiety. *Early Education & Development*, *24*(4), 446-467.

Vygotsky, L. (1978). Interaction between learning and development. *Readings on the Development of Children*, *23*(3), 34-41.

Wang, S. (2004). Online or paper: Does delivery affect results. *Administration Mode Comparability Study for Stanford Diagnostic Reading and Mathematics Tests*. Online paper. http://images.pearsonassessments.com/images/tmrs/tmrs_rg/OnlineorPaper.pdf

Wei, X., Lenz, K. B., & Blackorby, J. (2013). Math growth trajectories of students with disabilities: Disability category, gender, racial, and socioeconomic status differences from ages 7 to 17. *Remedial and Special Education*, *34*(3), 154-165.

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.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). Woodcock-Johnson III tests of achievement.

*Zippert, E. L., & Ramani, G. B. (2017). Parents' Estimations of Preschoolers' Number Skills Relate to at-Home Number-Related Activity Engagement. *Infant and Child Development*, *26*(2), e1968.

*Zippert, E. L., & Rittle-Johnson, B. (2018). The home math environment: More than numeracy. *Early Childhood Research Quarterly*.

BIOGRAPHICAL SKETCH

CURRICULUM VITAE

MIA CRISTINA DAUCOURT
Tallahassee, FL 32303

EDUCATION

08/2016 - Present **PhD in Developmental Psychology**
Florida State University, Tallahassee, FL
Graduate Adviser: Dr. Sara A. Hart
Overall GPA: 3.964

04/2016 **Bachelor of Science in Psychology**
Florida State University, Tallahassee, FL
Minor: Statistics
Overall GPA: 4.00, Summa cum Laude
Honors Thesis

07/2014 **Associate of Arts**
Tallahassee Community College, Tallahassee, FL
Overall GPA: 4.00

RESEARCH EXPERIENCE & POSITIONS

08/18 – Present **Graduate Research Assistant**
Florida Learning Disabilities Research Center: Engagement Core, Department of Psychology, Florida State University, Tallahassee, FL
Supervisor: Sara A. Hart, Ph.D.
Activities:

- Continued drafting monthly blog posts based on empirical papers to disseminate the FLDRC's empirical findings to the public
- Attended bi-weekly lab meetings to review and discuss recent empirical studies relating to achievement with the undergraduate research assistants in the lab to help them understand the methods and the state of the science in psychological achievement research
- Assisted in tracking personnel on the FLDRC grant, and created graphs and short write-ups to the FLDRC montly newsletter
- Helped coordinate a craft activity at the community event "Be My Neighbor Day" as a representative of FCRR to make bookworms with the children and families in attendance

- Underwent pedagogical training for 3 hours per week to become an undergraduate instructor for DEP3101 Child Psychology in Spring 2020, for which I created a full course syllabus and created and presented two 30-minute micro-lectures, as well as a 50-minute guest lectures with an average rating of 4 to 5 (out of 5) on all teaching characteristics, including clarity, engagement, presentation- and speaking-style.

09/17 – 08/18 **Graduate Research Assistant**
Florida Learning Disabilities Research Center: Engagement Core, Department of Psychology, Florida State University, Tallahassee, FL
Supervisor: Sara A. Hart, Ph.D.
Activities:
- Created a pre-registration for the recruitment of the new FLDRC National Twin Project on Achievement in Twins (NatPAT)
- Co-hosted a webinar on pre-registration and open science practices
- Drafted a monthly blog post based on empirical papers written with funds from the FLDRC grant
- Assisted in grant preparations with research and copy-editing
- Attended weekly meetings to discuss goals and duties for NatPAT and the Engagement Core as well as weekly lab meetings to discuss empirical research
- Trained DIS students on research practices, such as data entry and management and meta-analysis coding techniques
- Put together and transported packets of math assessments, prizes, pencils and other supplies in order to administer testing in Tallahassee elementary schools for a grant exploring children's math skills and math anxiety
- Administered assessments on early math skills and attitudes, like measurement and math anxiety, in two Tallahassee elementary schools to groups of 10-20 children at a time, following a standardized protocol

08/16 – 08/17 **Graduate Teaching Assistant**
Department of Psychology, Florida State University, Tallahassee, FL
Supervisor: Sara A. Hart, Ph.D. and Tiffany Hardy-Williams, Ph.D.
Activities:
- Guided 29 undergraduates through their first survey-based group research projects to be presented at an academic poster session at the end of the semester
- Held class for two hours every week and office hours for 1-3 hours per week, depending on student demand
- Taught students the scientific method and the important steps needed for conducting ethical, scientifically-sounds psychological research, including literature review, writing a research proposal, running

descriptive and inferential analyses using SPSS, drafting an informed consent and debriefing form, completing an IRB application for research with human subjects, and translating psychological measures into survey format using Qualtrics

- Lead lecture-based classroom instruction with weekly content in PowerPoint format, while balancing the presentation of new material with encouraging student engagement and answering questions
- Provided clear instructions for weekly lab assignments, including the provision of sample documents, SPSS code and accompanying screenshots
- Taught students all aspects of proper APA formatting
- Trained students to understand and interpret SPSS output
- Graded weekly individual and group assignments and pre-tests based on standard rubrics provided by the lecture professor
- Created a class website using the Canvas learning management system, including a folder for every class that included lecture slides, sample documents, SPSS data sets, and assignment submission links

08/16 – 08/17 **Undergraduate Advisor**
Department of Psychology, Florida State University, Tallahassee, FL
Supervisor: Kenneth Range and Tiffany Hardy-Williams, Ph.D.
Activities:

- Worked 5 hours per week in the Psychology undergraduate advising office providing advice, feedback, and sample application materials to undergraduates interested in pursuing a graduate degree or finding a research position
- Conducted online research and reached out to faculty members in order to find valuable resources, like listservs, for undergraduates looking for post-baccalaureate positions
- Provided written feedback on personal statements, resumes, and curricula vitae
- Helped students navigate academic websites to find graduate program information and communicate with faculty members and advisers
- Guided students through the application process, including etiquette for contacting faculty, the differences between research-based and skills-based programs, preparing for the GRE, building relationships with faculty members, and asking for reference letters

08/16 – 08/17 **Graduate Research Assistant**
Learning Disabilities Research Center, Florida State University, Tallahassee, FL
P.I.: Sara A. Hart, Ph.D.
Activities:

- Worked at least 20 hours per week conducting independent research for three papers in prep for publication

- Supervised and directed lab activities and duties for 4 DIS students and 1 work-study student, serving as a point of contact and mentor
- Applied for and attained IRB approval for new data collection on math anxiety and Common Core attitudes
- Coordinated and helped create Qualtrics surveys for new data collection on math anxiety and Common Core attitudes
- Coded for and created screens to be used by DIS students for data entry of Key Math scores for Project KIDS using FileMaker
- Managed data collection, troubleshooting any issues that arise with FileMaker
- Trained work-study student to code and manage the data screens and data in
- FileMaker and assist with other lab activities
- Attended weekly lab meetings to discuss research and plan for weekly duties to assign to DIS students
- Conducted statistical training workshop for undergraduate DIS students to learn basic SPSS analyses on February 20, 2017
- Individually mentored two DIS students that successfully presented independent research projects at the FSU Undergraduate Research Day on March 31, 2017

08/15 – 08/16 **Lab Coordinator,** Honors Thesis Research
Department of Psychology, Florida State University, Tallahassee, FL
Supervisor: Sara A. Hart, Ph.D.
Activities:

- Proposed original topic on the achievement effects of executive functions on response to intervention
- Spent at least 9 hours per week conducting independent research in order to compose a comprehensive prospectus for approval by the FSU Honors Department
- Applied for and obtained IRB approval for ethical research with kids
- Created original SAS code to analyze data from over 2,000 students and up to 50 different cognitive, behavioral, and other psychological measures
- Collaborated with Dr. Hart to start drafting the first paper using Project KIDS data to be submitted for publishing
- Coordinated undergraduate DIS student schedules and oversaw their projects
- Coded for and created screens to be used by DIS students for data entry of Key Math scores for Project KIDS in FileMaker
- Attended weekly lab meetings to discuss research and plan for the impending deadlines and weekly duties to assign to DIS students

01/15 – 08/15 **Research Assistant,** Independent Study with Honors
Department of Psychology, Florida State University, Tallahassee, FL

Supervisor: Sara A. Hart, Ph.D.
Activities:

- Proposed original topic on the achievement effects of social skills on response and problem behaviors on literacy gains over a school year
- Conducted logistic regression analyses and created graphical displays of data study results using original SAS code
- Created an original poster to present the results of the data study to undergraduates and faculty at FSU in a clear, concise manner
- Managed and organized data collected form online Qualtrics surveys using Excel and FileMaker, developing a familiarity and in-depth knowledge of all 3 programs
- Attended numerous recruitment events to procure contact information for new study participants for the FSU Center for Developmental Science family registry
- Presided over a phone recruitment team for the Hart lab that conducted screening interviews and collected important background information to maximize efficiency and effective placement of participants in corresponding studies
- Administered a full study protocol investigating the effects of screen reading on comprehension, which included administration of literacy and spatial reasoning assessments and use of Adobe and Flipbook 4-6 times per week
- Attended weekly lab meetings to discuss research and plan for the upcoming week of duties and deadlines

08/14 – 12/14 **Research Assistant,** Independent Study
Department of Psychology, Florida State University, Tallahassee, FL
Supervisor: Sara A. Hart, Ph.D.
Activities:

- Organized and managed data from over 3,000 Project KIDS participants using FileMaker, Excel and SAS
- Created original SAS code to calculate scores based on standardized scoring procedures for each scale and subscale included in the Project KIDS questionnaire
- Coded for conversion of assessment data into standard scores in FileMaker
- Entered hundreds of item-level responses to Woodcock Johnson-III Tests of Achievement subtests in order to build an electronic database that could be utilized for analysis
- Compiled mailings and other recruitment materials in order to maximize the number of participants and amount of data included in Project KIDS
- Attended weekly lab meetings to present and discuss empirical developmental research

08/14 – 01/16 **Research Assistant,** Independent Study
Department of Psychology, Florida State University, Tallahassee, FL, Anxiety and Behavioral Health Clinic

Supervisor: Norman B. Schmidt, Ph.D.
Activities:

- Conducted research on newly-developed computerized treatment targeting risk factors associated with suicide, mood, and anxiety symptoms funded by the Military Suicide Research Consortium
- Assisted in collection of psychophysiological data within a clinical population sample using EEG and peripheral (i.e.., EKG, SCR) measurement equipment
- Ran full 3-hour research protocol 1-3 times per week, which included 8 cognitive and emotional E-Prime tasks that measured effectiveness of cognitive bias modification training, from set up and data collection, to clean up and break down
- Pre-processed psychophysiological data with Neuroscan in real-time to verify the data collection was not impeded and data was viable for later use in analysis
- Trained graduate and undergraduate students in EEG protocols and methodology with close attention to detail to ensure conservation and proper use of laboratory equipment, including computers, EEG electrode bundles, and EEG machine
- Organized and maintained participant data utilizing Excel in order to create an electronic database that could be utilized for analysis and shared with colleagues in other geographic locations
- Conducted literature searches to determine the latest methods of analysis for certain kinds of data, like skin conductance response
- Attended monthly lab meetings to discuss the ongoing research and new trends in the field of neurophysiological research

PUBLICATIONS

Daucourt, M., Haughbrook, R., Taylor, J., & Hart, S.A. (2019, under review). The Nature and Nurture of the Association of EF and Reading with Math Performance. *Developmental Science*

Daucourt, M., Erbeli, F., Little, C.M., Haughbrook, R., & Hart, S.A. (2019). A Meta-Analytical Review of the Genetic and Environmental Correlations between Reading and Attention-Deficit Hyperactivity Disorder Symptoms and Reading and Math. *Scientific Studies of Reading*. doi: 10.1080/10888438.2019.1631827.

Daucourt, M., Taylor, J., & Hart, S.A. (2019, in prep). Gene x Socioeconomic Status Interaction on Decoding, Reading Comprehension, Writing, and Math, in an Economically-Diverse U.S. Twin Sample.

Daucourt, M., Napoli, A., Wood, S.G., Quinn, J., & Hart, S.A. (2019, in prep). The Home Math Environment and Math Achievement: A Meta-Analysis.

Barroso, C. Ganley, C.M., McGraw, A.L., Geer, E.A., **Daucourt, M.**, & Hart, S.A. (2019, in prep). A Meta-Analysis Investigating the Relation Between Math Anxiety & Math Achievement.

Daucourt, M., Little, C., Schatschneider, C., Petscher, Y., Haughbrook, R., & Barroso, C. (2019, in prep). The Black-White Achievement Gap in Reading: A Linear Quantile Mixed Modelling Approach. *Journal of Educational Psychology*

Daucourt, M., Schatschneider, C., Connor, C., Al Otaiba, S., & Hart, S.A. (2018). Updating Working Memory, Inhibition, and Shifting Predict Reading Disability Symptoms in a Hybrid Model: Project KIDS. *Frontiers in Psychology*

Hart, S.A., **Daucourt, M.**, & Ganley, C.M. (2017). Individual Differences Related to College Students' Couse Performance in Calculus II. *Journal of Learning Analytics*, 4(2), 129-153.

POSTERS & PRESENTATIONS

Daucourt, M., & Edwards, A. A. (2018). Classification and Regression Tree (CART) Analysis of Early Reading Outcomes. Poster presented at the Machine Learning Conference (Tallahassee, FL). April 20, 2019.

Edwards, A. A., & **Daucourt, M.** (2018). Classification of Response to Intervention for Early Reading. Poster presented at the Machine Learning Conference (Tallahassee, FL). April 20, 2019.

Daucourt, M., Napoli, A., Wood, S. G., & Hart, S. A. (2019). The Home Math Environment and Math Achievement: A Meta-Analysis. Poster presented at Graduate Research Day (Tallahassee, FL). April 12, 2019.

Barroso, C. Ganley, C.M., McGraw, A.L., Geer, E.A., **Daucourt, M.**, & Hart, S.A. (2019). A Meta-Analysis Investigating the Relation Between Math Anxiety & Math Achievement. Poster presented at Graduate Research Day (Tallahassee, FL). April 12, 2019.

Daucourt, M., Erbeli, F., Little, C. M., Haughbrook, R., & Hart, S.A. (2018). Genetic and Environmental Correlations between Reading and Attention-Deficit Hyperactivity Disorder Symptoms and Reading and Math: A Meta-Analytical Review of Twin Studies. Poster presented at the Learning Disabilities Research Center Consortium (Houston, Texas). November 29, 2018.

Daucourt, M., Haughbrook, R., Taylor, J., & Hart, S.A. (2018). Gene x Socioeconomic Status Interaction on Decoding, Reading Comprehension, Writing, and Math, in an Economically-Diverse U.S. Twin Sample. Symposium talk presented at Society for the

Scientific Study of Reading 25th Annual Meeting (Brighton, England). July 18-21, 2018.

Daucourt, M., Haughbrook, R., Taylor, J., & Hart, S.A. (2018). The Nature and Nurture of the Association of EF and Reading with Math Performance. Talk to be presented at FSU Graduate Research Day (Tallahassee, FL). April 27, 2018.

Daucourt, M., Haughbrook, R., Taylor, J., & Hart, S.A. (2017). The Nature and Nurture of the Association of EF and Reading with Math Performance. Poster presented at Vanderbilt University Math Cognition Conference (Nashville, TN). May 15-16, 2017.

Daucourt, M., Little, C., Schatschneider, C., Petscher, Y., Haughbrook, R., Barroso, C., & Hart, S.A. (2017). The Black-White Achievement Gap: A Linear Quantile Mixed Model Analysis. Poster presented at FSU Graduate Research Day (Tallahassee, FL). April 21, 2017.

Daucourt, M., Schatschneider, C., Connor, C.M., Al Otaiba, S., & Hart, S.A. (2017). Project KIDS: EF Predicts "Responders" and Non-responders" in a Constellation Model of Reading Disability. Poster presented at SRCD Biennial Meeting (Austin, TX). March 6-8, 2017.

Daucourt, M. & Hart, S.A. (2017). The Black White Achievement Gap in Reading. Talk presented at FSU Cognitive Brown Bag Series (Tallahassee, FL). March 10, 2017.

Daucourt, M. & Hart, S.A. (2016). Executive Functioning and Response to Intervention: Project KIDS. Poster presented at 28th Annual APS Convention (Chicago, IL). May 26-29, 2016.

Daucourt, M., & Hart, S.A. (2016). Executive Functioning and Response to Intervention: Project KIDS. Howard D. Baker talk and poster presented at FSU Undergraduate Research Day (Tallahassee, FL). April 1, 2016.

Daucourt, M., & Hart, S.A. (2015). How Problem Behaviors Affect Residualized Literacy Gains Over a School Year. Poster presented at FSU Undergraduate Research Day (Tallahassee, FL). April 10, 2015.

VOLUNTEER EXPERIENCE

02/2017 – Present **MATH PALS**

Sealey Elementary School, Springwood Elementary School, Tallahassee, FL

Activities:

- Met weekly for one hour on site with assigned mentee for individual instruction in elementary math

- Taught a new literature-based math lesson every week on a range of topics, including multiplication, division, arrays, measurement, and many others
- Answered mentee's questions through verbal explanation and visual demonstration
- Covered a list of topic-related math vocabulary words every week to support the new concepts learned
- Demonstrated math concepts using props and real-life examples in order to apply newly-learned math material
- Created a rapport as an instructor and friend with mentees, Elijah, Mikkel, and Darnelle

08/2015 – 07/2016 **SHIFT SUPERVISOR AND MENTOR**

Big Bend Crisis Hotline, Tallahassee, FL

Activities:

- Completed 20 hours of additional training for certification
- Supervised newly-trained volunteers on the phones to provide back-up, feedback and evaluation on their phone room calls for 6-12 hours per week
- Served as a mentor for new volunteers to answer questions and ease their transition into the agency
- Coached volunteers on how to handle difficult calls and follow agency procedures even in novel situations

01/2015 - 07/2016 **FLORIDA HIV & AIDS HOTLINE**

Big Bend Crisis Hotline, Tallahassee, FL

Activities:

- Completed 30 hours of additional training for certification
- Provided short-term crisis counseling, information and referrals for HIV/AIDS, STDs, and other communicable diseases for all of Florida 6-12 hours per week
- Connected callers with financial support resources in order to help them attain subsidies for medication and other costs incurred due to HIV diagnosis
- Informed callers about the risks associated with STDs, including their window periods, treatments, and how to get tested
- Formed a short-term therapeutic relationship with callers to provide them with a forum to express their feelings

08/2014 - 07/2016 **HELPLINE 211 & LIFELINE**

Big Bend Crisis Hotline, Tallahassee, FL

Activities:

- Completed a semester-long, 5-10 hours per week training program based on Carl Rogers' RIDE model that required a 90% or higher on a written final exam and phone call training for certification
- Provided short-term crisis counseling, information and referrals for 8 counties of the Big Bend service area and the United Way for 20-40 hours per month
- Established a therapeutic relationship with the caller to build rapport and make the caller feel comfortable sharing personal information
- Identified the callers' specific needs and feelings within the first two minutes of the phone exchange in order to stay focused and help the caller come up with solutions
- Explored alternative solutions in a non-directive manner to increase the callers' resourcefulness and ability to cope with crises

HONORS, SPECIAL AWARDS, & MEMBERSHIPS

HONORS AND AWARDS

2018 Congress of Graduate Students International Presentation Grant ($500)
2017 Jane M. West Research Fellowship ($500)
2017 Russell and Eugenia Morcom GRD Excellence Award: Best Poster in Developmental
2017 Congress of Graduate Students Attendance Grant ($100)
2016 Mark DeGraff & Lulu Hamilton DeGraff Research Scholarship ($1,500)
2016 Legacy Fellowship ($10,000 annually for 5 years of Ph.D. program)
2016 Mark DeGraff & Lulu Hamilton DeGraff Research Scholarship ($1,500)
2016 Mae Hamptom Watt Presidential Scholarship in Psychology: Excellence in Research Award ($1,000)
2016 Howard D. Baker Undergraduate Research Award, 2nd place ($200)
2016 Student Council for Undergraduate Research and Creativity Travel Grant ($500)

MEMBERSHIPS

2019 Diverse Psychology Graduate Student Organization
2017 Psychology Department Representative for Graduate Student Advisory Council
2016 Society for Research in Child Development
2016 Association for Psychological Science
2015 Phi Beta Kappa Honor Society
2015 Golden Key International Honor Society
2015 National Alliance of Mental Illness
2015 Psi Chi Honor Society

CERTIFICATIONS

04/2016 Certification in SAS Programming & Data Analysis

WORKSHOPS

04/2017	Attendee, Workshop: “International Workshop on Statistical Genetic Methods for Human Complex Traits”. University of Colorado Boulder, Boulder, CO.
02/2016	Coordinator, Workshop: “Introduction to SPSS”. Florida State University, Tallahassee, FL.
06/2016	Attendee, Workshop: “Quantile Regression”. Florida State University, Tallahassee, FL.
06/2016	Attendee, Workshop: “Getting Acquainted with R”. Florida State University, Tallahassee, FL.
11/2015	Attendee, Workshop: “Behavioral Genetics Boot Camp”. Florida State University, Tallahassee, FL.

COURSES COMPLETED TO ADVANCE METHODOLOGICAL TRAINING

Spring 2019	Advanced Quantitative Methods
Fall 2018	Introduction to Structural Equation Modeling
Fall 2017	Meta-Analysis
Spring 2016	Regression
Spring 2016	Hierarchical Linear Modeling
Fall 2016	Analysis of Variance