THREE RESEARCH TOPICS IN EDUCATION: (1) ASSOCIATIONS BETWEEN APPROACHES TO LEARNING AND ACADEMIC ACHIEVEMENT; (2) A META-ANALYTIC REVIEW ON APPROACHES TO LEARNING AND ACADEMIC ACHIEVEMENT; (3) POWER ANALYSIS IN META-ANALYSIS: A THREE-LEVEL MODEL

By

Bixi Zhang

A DISSERATATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Measurement and Quantitative Methods—Doctor of Philosophy

2021

ABSTRACT

THREE RESEARCH TOPICS IN EDUCATION: (1) ASSOCIATIONS BETWEEN APPROACHES TO LEARNING AND ACADEMIC ACHIEVEMENT; (2) A META-ANALYTIC REVIEW ON APPROACHES TO LEARNING AND ACADEMIC ACHIEVEMENT; (3) POWER ANALYSIS IN META-ANALYSIS: A THREE-LEVEL MODEL

By

Bixi Zhang

This dissertation is a three-piece dissertation, including two empirical research (Chapter 1 and Chapter 2) and a methodological improvement of prior work (Chapter 3), to address issues of the effects of approaches to learning on academic achievement in childhood and power analysis for a three-level model in meta-analysis.

Approaches to learning as a key domain of school readiness has shown significant effects on student academic achievement. The study in Chapter 1 was designed to examine the potential moderation effects of problem behaviors on the association between approaches to learning and academic achievement (mathematics, reading, and science) in early grades using a recent nationwide longitudinal dataset (ECLS-K:2011). The correlated random effects estimation was applied to deal with the omitted variable issue. At the same time, the estimation method was allowed to compute the impacts of important time-constant variables (e.g., socioeconomic status) on academic achievement. The results indicated non-significant moderation effects of problem behaviors on the relations between approaches to learning and academic outcomes. However, the main effects of approaches to learning were significant associated with academic achievement. Complete data analysis and bootstrap with multiple imputation were conducted in the research to address the missing data issue in large-scale assessments. Similar results were shown in the two approaches, which demonstrated robust findings in the study.

To better understand the general relations of approaches to learning on academic performance in childhood in recent years, the study in Chapter 2 conducted a systematic review employing meta-analytic methodology to combine and summarize the results from empirical quasi-experimental studies. The study filled the literature gap and extended the theory to understand the relations between approaches to learning and achievement. The results indicated medium effect sizes of the relations on approaches to learning and concurrent/future achievement (reading and mathematics). The effects on reading achievement were slightly larger than the effects on mathematics achievement. The single timepoint evidence showed stronger effects compared with longitudinal designs. In conclusion, the meta-analysis emphasized the positive and significant effects of approaches to learning on academic achievement in childhood.

The methodological improvement in Chapter 3 aimed to address the potential biased power statistics when introducing group dependence in meta-analysis. The study extended the prior work about power analysis for two-level random-effects models to three-level models in the univariate case. The three-level model assumed research teams/labs at the third level. A three-level random-effects model provides more accurate estimates of power under the assumption that variability between research teams is not negligible. Each model in the study was followed by an illustrated example to show how to calculate the power. A simulation study provided evidence about how group-level heterogeneity affected statistical power in meta-analysis in the three-level model. The present study introduced more complicated data structures in meta-analysis and provided the power measures in advanced meta-analytic models.

ACKNOWLEDGEMENTS

I would first like to express my deepest appreciation to my advisor and committee chair, Dr. Spyros Konstantopoulos, for your tireless and unlimited guidance and support over the past five years in my doctoral studies. Your professional mentorship, knowledgeable experiences, and rigorous attitude deeply inspire me and help me become a good scholar in quantitative methods in education. I am extremely grateful for all your encouragement and help. The completion of my dissertation would not have been possible without the nurturing of you.

I would also like to extend my deepest gratitude to my committee members: Dr. Jeffery Wooldridge, Dr. Kenneth Frank, and Dr. Yuehua Cui, for your expertise and continuous support. Thank you for taking an interest in my work. Your assistance and professional suggestions encourage me to think deeper and help me accomplish the practicum project and dissertation. I would like to acknowledge the Graduate School Dissertation Completion Fellowship for financially supporting the dissertation.

I would like to extend my sincere thanks to my grandparents, Tongjin Zhang and Qiying Dai, for giving me a happy and carefree childhood and educating me to become a brave girl. I am grateful to my parents, Li Zhang and Ran Xu, for your unlimited love and unconditional support. Thank you for nurturing and educating me to be an independent woman. I would like to thank my husband, Dr. Anqi Chen, for your unlimited love and emotional support especially during the pandemic years.

Thank you should also go to my colleagues and friends, Dr. Xiaowan Zhang, Dr. Qinyun Lin, Dr. Zixi Chen, Dr. Gu Zheng, Dr. Sohyeon Bae, Xuran Wang, Jiawei Li, Jiachen Liu, Heqiao Wang and Zhonghao Wang, in Michigan State University. I appreciate your professional suggestions to my work and company and support in my life. Last but not least, I must also thank my best friends in my life, Yixin Fu, Xiaomeng Wang, and Wenke Bian. Thank you for your company and continuous emotional support over ten years.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 APPROACHES TO LEARNING AND ACADEMIC ACHIEVEMENT IN ELEMENTARY SCHOOL: TESTING MODERATING ROLES OF PROBLEM BEHAVIOURS

Introduction

When children start their school years, a fundamental aspiration for schoolers is to prepare school readiness and later success. Schools often focus on children's cognitive development (e.g., academic achievement), especially elementary schools, however, approaches towards learning is another key component that can help students succeed in schools. In 2019, the U.S. Department of Health and Human Services published a framework of school readiness, Head Start Early Learning Outcomes (revised version), which was designed to represent the continuum of learning in early childhood. Approaches to Learning (ATL) was designed as a core domain in this framework, which refers to the skills and behaviors that children use to engage in learning activities. ATL is essential for children in early grades which was emphasized by prior theories and frameworks. Kagan et al. (1995) stated that "ATL frame the child's entire being and are at the core of social/emotional and cognitive interactions" (p.28). U.S. Department of Health and Human Services (2019) claims that supporting skills and behaviors that children use to engage in learning could help them develop well in all domains and contributes to school success directly. The improvement and mastery of learning approaches is associated both with students' school transition, school performance as well as social-emotional outcomes (McClelland & Morrison, 2003; Atkins-Burnett, 2007). For instance, children who were at kindergarten with inadequate learning-related skills were at a greater risk in elementary school and lower academic performance (McClelland et al., 2006). ATL was highly related to kindergarten retention (Hong & Raudenbush, 2005).

There is an increasing interest in investigating the importance of the effects of learning-related behaviors on academic outcome in childhood (McClelland & Morrison, 2003). Considering the potential influence of ATL on student academic achievement, the current study examines the association between ATL and reading, mathematics and science achievement from kindergarten to fifth grade, in order to understand generally how ATL have impacts on those cognitive achievement. In addition, besides ATL, problem behaviors (PB), which is another type of behaviors that could influence school success significantly, were tested by previous studies. For example, Malecki et al. (2002) showed that problem behaviors were negatively associated with academic achievement in elementary school years. Because ATL relates to the behaviors during learning tasks, problem behaviors could possibly change the association of ATL and student achievement. However, few longitudinal studies using the evidence from elementary school test the effect of problem behaviors in the relations between ATL and achievement. The present study aims to investigate the impact of problem behaviors on the relationship between ATL and achievement by using a recent national longitudinal assessment.

Literature Review

History of Approaches to Learning

Back to 1970s, Anderson and Messick (1974) created twenty-nine statements which represented a group of theory-guided components systematically in social competency among young children. The statements showed some relevant components of learning dispositions/approaches which were defined later. In 1980s, Katz (1985) emphasized the importance of learning disposition as one of the three goals for early childhood education. Leaning dispositions were defined by Katz as “relatively enduring ‘habits of mind’ or the characteristic ways of responding to experience across types of situations” (p.1). Those definitions helped later researchers to generate a group of components which represent approaches toward learning. In 1990s, the Nation Education Goals Panel (NEGP) defined ATL as “inclinations, dispositions, or styles that reflects the myriad ways that children become involved in learning.” (Kagan et al., 1995, p.4). Because of the lack of ATL instruments matching with definitions from NEGP and Katz (Meisels et al., 1996), Atkins-Burnett developed a rating scale for ATL which covered the considerations from the definitions and met the needs of the Early Childhood Longitudinal Study (ECLS) in the United States in the late 1990s.

Moving to 21st century, to get a comprehensive understanding about the approaches and behaviors that students show in the learning process, the Habit of Mind framework was developed. The framework indicated that learning dispositions, such as persisting and managing impulsivity, could facilitate children’s thinking and learning, at the same time, building a thoughtful classroom environment (Costa & Kallick, 2008). Recent framework showed that academic enablers contributed to achievement (DiPerna et al., 2002; DiPerna & Elliott, 2002). The academic enablers could help to improve the academic skill so that improve the academic achievement (DiPerna, 2006). The learning-related behaviors from academic enablers are often referred to ATL (Anthony et al., 2014). Specifically, ATL includes students’ behaviors, strategies and attitudes in learning contexts or educational tasks with components such as self-regulation, persistence and attentiveness (Li-Grining et al., 2010; McWayne et al., 2004).

Approaches to Learning and Academic Achievement

Previous empirical research has shown that ATL reliably predicts student achievement. In particular, McWayne et al. (2004) found that ATL significantly impacted performance on the kindergarten version of the Early Screening Inventory, which is a test that reflects children's early academic success. ATL significantly predicted gains in science and mathematics among a group of preschoolers from low-income families (Bustamante et al., 2017). Other studies also established that several important components from ATL definitions and frameworks were closely related to academic outcomes in childhood. For instance, behavioral self-regulation and executive function were strongly related to academic growth in mathematics, literacy and vocabulary in prekindergarten and kindergarten (McClelland et al., 2014). Children's emotional regulation and behavioral self-regulation were also found to be positively related with academic achievement in kindergarten and early grades (Howse et al., 2003; McClelland & Cameron, 2011). Furthermore, preschooler's persistence can predict academic achievement in kindergarten to a greater degree than demographic variables and cognitive-linguistic skills (Mokrova et al., 2013).

Research has also indicated that higher levels of learning-related skills were linked to higher reading and mathematics achievement, however, learning-related skills appeared to have a stronger effect on children's test scores in kindergarten and early primary school grades than in later grades (McClelland et al., 2006). Along the same lines, ATL at kindergarten entry was a reliable predictor of growth in reading and mathematics achievement from kindergarten through fifth grade (Li-Grining et al., 2010). Moreover, students with higher teacher report ATL ratings had higher achievement in reading and mathematics in early grades (Hong & Yu, 2007). Bodovski and Farkas (2007) found that ATL measures had a strong association with mathematics growth from kindergarten to third grade. Recent research showed that a significant association between ATL and mathematics achievement from kindergarten to second grades (Ribner, 2020).

Researchers have also examined the relationship between ATL and student academic achievement in higher education (Duff et al., 2004). For example, self-regulated learning and motivation have been linked to academic achievement among undergraduate students (Mega et al., 2014). In addition, a recent study showed that improvements in self-regulation were related with higher levels of achievement for college freshmen (Wibrowski et al., 2017).

Trainings and interventions for learning approaches have been utilized to help children decrease the risk of various problems they may face and improve their learning (Zin, 2004). Evaluation results indicated that the improvement of ATL would facilitate student academic performance. Preschoolers who participated in an eight-week self-regulation intervention got higher academic achievement gain compared with the control group (Schmitt et al., 2015). Perels et al. (2009) showed that trainings on self-regulative strategies significantly improved student mathematics achievement. Preventive curricula about enhancing attention skills could be helpful to improve young children's future academic success (Rhoades et al., 2011).

The Role of Problem Behaviors in the Relation

Besides ATL, children's problem behaviors strongly influence students' school success in early grades as well. Problem behaviors are stable and could affect later social, emotional and academic functioning (Campbell, 1995). Previous studies have shown consistently that problem behaviors are negatively predictive of academic achievement in elementary school (Malecki et al., 2002; Nelson et al., 2004; Algozzine et al., 2011). In addition, problem behaviors potentially reflect learning disabilities and emotional disturbance in early grades (Algozzine et al., 2011). Schaefer and colleagues (2004) found that learning-related behaviors were negatively associated with problem behaviors in kindergarten. Early problem behaviors predicted lower academic outcomes and lower rating of approaches to learning, such as attention and persistence (Bulotsky-Shearer et al., 2011). Razza and colleagues (2015) found that early problem behaviors in kindergarten could influence the relationship between ATL at age 5 and social competence in later. However, rare study tested the relationship among ATL, problem behaviors and academic achievement before. Therefore, the current study hypothesizes that problem behaviors possibly have moderation effect on the relationship between ATL and academic achievement.

The hypothesized role of problem behaviors in the relationship is called a moderator. A moderator is a qualitative or quantitative variable that affects the direction and/or strength of the relation between an independent or predictor variable and a dependent or criterion variable (Baron & Kenny, 1986). The moderation effect of problem behaviors could also be called the interaction effect between ATL and problem behaviors in the statistical model. In the current study, the term "interaction effect" and "moderation effect" would be used interchangeably.

Present Study

The present study moved forward to examine the interaction effect of ATL and problem behaviors on academic achievement using a nationwide longitudinal dataset. In other words, the study investigated a more complex mechanism among ATL, problem behaviors and academic performance. The results of the study shed lighter about how problem behaviors moderated the association between learning-related skills and academic outcomes in elementary school from a longitudinal perspective. Additionally, most prior studies applied growth models or traditional regression models to examine the association between ATL and academic achievement (Li-Grining et al., 2010; McClelland et al., 2006). However, these methods may potentially suffer from omitted variable bias. To address this caveat, this study introduced a linear unobserved effects panel data model, which controlled well for unobserved student-level time-constant effects. Specifically, the correlated random-effects (CRE) model (Wooldridge, 2005; Wooldridge, 2010), was proposed first by Mundlak (1978). It can eliminate unobserved individual time-constant variables effects to get unbiased estimates, meanwhile report estimates of observed time-constant variables (Wooldridge, 2010). Thus, the CRE estimation is a good fit to answer the research questions because the data at hand are panel (longitudinal), and in such cases, this estimation reduces selection bias due to time-constant unobserved variables and could provide the estimates of relevant time-constant covariates. Therefore, the purpose of this study was to investigate three research questions:

(1) Is there any interaction effect of ATL and problem behaviors (moderation effect) on achievement (reading, mathematics, science) in elementary school grades, K-5?

(2) If no moderation effect of problem behaviors exist, does ATL have an effect on student achievement from kindergarten to fifth grade controlling for problem behaviors?

(3) Are any time-constant covariates significant?

The study had following novelties. First, it extended previous works on the relationship between ATL and achievement from early childhood to the whole childhood by conducting a longitudinal analysis and introduced problem behaviors as potential moderators in the association between ATL and achievement. Second, the study used appropriate statistical methods that control for all unobserved individual-level time-constant effects, at the same time, provided important observed time-constant effects on achievement. Third, by using the most recent nationwide longitudinal assessment in education under current demographic environment, the results were timing and convincing. Fourth, from a practical perspective, the finding could help educators and teachers understand students' behaviors better, thus taking quick instructions and interventions could possibly facilitate students' learning and school success.

Methods

Data Sources

The Early Childhood Longitudinal Study, Kindergarten Class of 2010-11 (ECLS-K:2011) is a large-scale longitudinal educational assessment survey supported by the National Center for Education Statistics (NCES), under the U.S. Department of Education. The assessment traced students for six years from kindergarten (2010) to fifth grade (2016). ECLS-K:2011 is a more recent dataset that focuses on academic achievement in the 21st century and provides researchers with an opportunity to analyze data that includes the information of the latest school developments and effects. Compared with the previous round of ECLS-K (ECLS-K:1998), policies in education and demographic environment have changed significantly after a decade. For instance, the policy of No Child Left Behind has been passed and children have broader choices of schools (NCES, n.d.). In addition, ECLS-K:1998 did not collect information from the second-grade year and the fourth-grade year. ECLS-K:2011 shows high reliability and validity of its measurement. The information could get from the User's Manual for the ECLS-K:2011 Kindergarten-Fourth Grade Data File and Electronic Codebook, Public Version, which was written and reported by Tourangeau and colleagues (2018). The study could provide more convening evidence by using ECLS-K:2011 dataset. All measures used in the study were from spring semester assessment (kindergarten to fifth grade). Table A.1 in Appendix A shows the name and description of the variables used in the study from ECLS-K:2011.

Outcomes

Children's reading, mathematics and science IRT scale scores were used as dependent variables in the analysis. Reading test specifications include basic reading skills, vocabulary, comprehension, mathematics test specifications include number properties and operations, measurement, geometry, data analysis and probability, and algebra, and science test specifications include scientific inquiry, life science, physical science and earth/space science. The reliability estimates were over 0.90 for each round of reading, mathematics and science assessments. The high validity of test scores has been verified by a review of standards from the nation and states, and the frameworks of tests were developed by referring other national assessment in education (i.e., National Assessment of Educational Progress) (Tourangeau et al., 2018).

Instruments of Approaches to Learning and Problem Behaviors

The study used teacher-report ATL scores which were constructed by ECLS-K:2011. In particular, ECLS-K:2011 created a composite score of ATL that consists of seven items about students' learning-related behaviors: keeps belongings organized; shows eagerness to learn new things; works independently; easily adapts to changes in routine; persists in completing tasks; pays attention well and follows classroom rules. A four-point Likert scale (from 1 to 4) was used for each item to rate students' learning behaviors from never to very often. Higher scale scores indicate that the child exhibited positive learning behaviors more often according to the teacher (Tourangeau et al., 2018). The Approaches to Learning Scale were developed specifically for ECLS-K studies. It used the same frequency scale of social skill items, which were adapted from the Social Skills Rating System (SSRS) by Pearson. The ATL scale has a high internal consistency reliability (0.91) for each round of assessment in ECLS-K:2011, which was reported in the psychometrics reports (Najarian et al., 2018a; Najarian et al., 2018b; Najarian et al., 2019). ECLS-K:2011 statistical/psychometrics team computed a mean score when the respondent provided a rating on at least four of the seven items. Therefore, the ATL composite score could be treated as a continuous variable in the analysis.

The study used teacher-report problem behaviors scores constructed by ECLS-K:2011. The items were adapted from the Social Skills Rating System (SSRS) by Pearson. The problem behaviors scales include two scales, externalizing and internalizing problem behaviors. The six-item externalizing behaviors scale measured the frequency with which a child argues, fights, gets angry, acts impulsively, disturbs ongoing activities, child's tendency to talk at times when the child was not supposed to be talking. The four-item internalizing behaviors scale measured the extent that the child exhibits anxiety, loneliness, low self-esteem, and sadness. A four-point Likert scale (from 1 to 4) was used for each item to rate students' problem behaviors from never to very often (Tourangeau et al., 2018). Both problem behaviors scales have high internal consistency reliability for each round of assessment in ECLS-K:2011 in spring semester. ECLS-K:2011 statistical/psychometrics team computed two mean score when the respondent provided a rating on at least four of the six items from externalizing problem behaviors scale and at least three of the fourth items from internalizing problem behaviors scale. Thus, two composite scores of externalizing and internalizing problem behaviors used in the study could be treated as a continuous variable.

Time-Varying and Time-Constant Covariates

The study selected proper time-varying variables as control variables in the model. These variables include teacher experience, school enrollment and school socioeconomic status (school SES). Finally, whether students changed schools or not was also included as a covariate in the statistical analysis. Because the study used teacher-report items as ATL and problem behaviors measures in the model, controlling teacher’s experience might be beneficial to avoid potential rating bias. The teacher’s experience was a continuous covariate in the model. School SES was represented by the variable that showed the percent of students eligible for free or reduced-price lunch in school. Due to the limitation of the dataset, school enrollment and school SES were ordinal variables in the dataset. The study recoded them as continuous variables using midpoints of the initial categories. Changing school or not was a binary variable in the model.

The study also selected proper observed time-constant variables to investigate whether these variables were the correlates of student academic achievement. These variables included age, socioeconomic status (SES), speaking non-English at home, gender, and race. Four race dummies (Black students, Hispanic students, Asian student, and other) were created to examine race differences in achievement (White students being the reference group). The interaction term between ATL and SES was included in the model to control the potential influence.

Missing Data Issue

Because the present study used a secondary large-scale longitudinal dataset, missing data issue potentially existed in the analysis. The study conducted a complete data analysis first to get a result from students who continuously provided data from kindergarten to fifth grade. The study also conducted resampling methods to deal with potential missing values. Bootstrap and multiple imputation strategy was applied in the study. Bootstrap is a computer-based simulation method and could reduce the bias and prediction error to achieve high statistical accuracy (Efron & Tibshirani, 1986). Multiple imputation is shown as a convenient and popular paradigm in the analysis of missing data (Schafer, 1999). The combination of bootstrap and multiple imputation is possible to deal with missing data issue and get robust results. Such strategy has been discussed and used in the previous studies (Comulada, 2015; Schomaker & Heumann, 2018). In detail, the study resampled the incomplete data with missing, then conducted multiple imputation with five times to get imputed data points for the dataset. The next step was to estimate the coefficients by using the same model in the complete data analysis. The analysis did bootstrap 1000 times to get more robust results. Every time the study resampled the data and imputed the missing data would give standard errors that account for the imputation. Therefore, the study would get the regression coefficients and the bootstrap standard errors.

Participants

The complete data analysis included 5735 students who had all data information in the six time periods. The bootstrap with multiple imputation analysis included over 10000 students (10702) who had part of information from the dataset.

Statistical Analysis

The present study applied a linear unobserved effects panel data model, which is called the correlated random effects (CRE) model, to find the potential relationships among ATL, problem behaviors and academic achievement in early grades. CRE approach allows us to include time-constant variables and simultaneously delivers the fixed effects estimates of the time-varying variables. The CRE model produced the coefficients of interest (i.e., ATL, problem behaviors, and the interaction between ATL and problem behaviors) where the potential impact of the time-constant confounding variables was removed (under the assumption that the fixed effects are linear). To conduct this analysis, the study used the panel data from six time periods (kindergarten to fifth grade).

Introduction to Correlated Random Effects Model

A linear unobserved effects panel data model is displayed in Equation (1.1). It presents a general equation of a linear unobserved effects panel data model with time-varying measure of the interested variables ($\boldsymbol{x_{it}}$), time-varying covariates ($\boldsymbol{w_{it}}$), time dummies ($\boldsymbol{D_t}$), and individual-level unobserved time-constant effect (c_i). The model could be treated as a fixed effects model in panel data analysis. Traditional fixed effects model could get the coefficients of time-varying variables. However, the estimations of observed time-constant variables are unavailable. The main idea of CRE model is to use time-constant and time-average variables to model the unobserved individual effect. An example of a CRE model is shown in equations as:

$$y_{it} = \boldsymbol{x_{it}\beta} + \boldsymbol{w_{it}\gamma} + \boldsymbol{D_t\delta} + c_i + u_{it}, \tag{1.1}$$

$$c_i = \psi + \boldsymbol{z_i\eta} + \bar{\mathbf{x}}_{\mathbf{i}}\boldsymbol{\xi} + \boldsymbol{\bar{w}_i\zeta} + a_i. \tag{1.2}$$

Equation (1.2) separates the individual-level unobserved time-constant effect into several parts, including an intercept (ψ), time average of measure the interested variable ($\overline{\boldsymbol{x}}_{\boldsymbol{i}}$), time-constant covariates ($\boldsymbol{z_i}$), and time averages of time-varying covariates ($\overline{\boldsymbol{w}}_{\boldsymbol{i}}$). Therefore, Equation (1.2) allows c_i to be correlated with the time-varying variables through its average levels over time (where ψ is the intercept and a_i is the error term) (see Wooldridge, 2010). To note that variables and parameters are in boldface indicate vectors in the equations.

Thus, replacing c_i in Equation (1.1) by using the model in Equation (1.2) could get the CRE estimating equation:

$$y_{it} = \boldsymbol{x_{it}\beta} + \boldsymbol{w_{it}\gamma} + \boldsymbol{D_t\delta} + \boldsymbol{z_i\eta} + \psi + \overline{\boldsymbol{x}}_{\boldsymbol{i}}\boldsymbol{\xi} + \overline{\boldsymbol{w}}_{\boldsymbol{i}}\boldsymbol{\zeta} + v_{it}, \tag{1.3}$$

where $v_{it} = a_i + u_{it}$ are the composite error at time period *t*. The error term is a sum of two parts error from Equation (1.1) and (1.2). From Equation (1.3), the CRE model allows us to get the estimates of the interested independent variables ($\boldsymbol{x_{it}}$) and time-constant variables ($\boldsymbol{z_i}$) from a single estimation model.

Traditional models in panel data analysis are the fixed-effects and random-effects estimation. The fixed-effects model is highly used in research because it allows the correlations between the unobserved heterogeneity and time-varying predictors (Wooldridge, 2005). However, these two approaches could not compute time-constant estimators. The CRE model has its advantages in getting the estimations of time-constant effects. At the same time, the CRE model can obtain the same estimators from the fixed-effects model. Thus, it increases flexibility in a straightforward model with a decomposition of within and between effects and combines advantages of fixed effects and random effects estimation (Schunck, 2013).

Correlated Random Effects Model with Interactions

The present study aimed to investigate the relationship between ATL and academic approaches to learning, and the potential moderation effects of problem behaviors in the relation. Univariate analyses were conducted in the study, which means the study had three similar models with different academic outcomes. The interaction terms (testing moderation effects) in the model were the products of time-demeaned variables. Thus, the interaction terms represented the within-unit association between interested variables (ATL) and outcome (academic achievement) with the intra-unit variation of the moderators (problem behaviors). These terms could control the effect of heterogeneity (Giesselmann & Schmidt-Catran, 2020). And it could avoid the potential multicollinearity of the main variables. Therefore, the CRE model with interaction terms could be written as:

$$y_{it} = x_{it}\boldsymbol{\beta} + (x_{it} - \bar{x}_{i.}) \times (\boldsymbol{w_{it}} - \boldsymbol{\bar{w}_{i.}})\boldsymbol{\lambda} + \boldsymbol{w_{it}}\boldsymbol{\gamma} + \boldsymbol{n_{it}}\boldsymbol{\varphi} + \boldsymbol{D_t}\boldsymbol{\delta} + \boldsymbol{z_i}\boldsymbol{\eta} + \psi + \bar{x}_{i.}\xi$$
$$+ \overline{(x_{it} - \bar{x}_{i.}) \times (\boldsymbol{w_{it}} - \boldsymbol{\bar{w}_{i.}})}\,\boldsymbol{\rho} + \boldsymbol{\bar{w}_{i.}}\boldsymbol{\zeta} + \boldsymbol{\bar{n}_{i.}}\boldsymbol{\phi} + v_{it}\,, \qquad (1.4)$$

where *t = K spring, 1st spring, …, 5th spring (six time periods),*

y_{it} *is the reading/mathematics/science IRT score in time t for individual i,*

x_{it} *is the ATL measure in time t for individual i,*

$\boldsymbol{w_{it}}$ *are problem behaviors (externalizing and internalizing) measures in time t for individual i,*

$\boldsymbol{n_{it}}$ *are proper time-varying covariates in time t for individual i,*

$\boldsymbol{z_i}$ *include proper observed time-constant covariates for individual i,*

$\boldsymbol{D_t}$ *are time dummies (controlling for time/grade effects),*

$\boldsymbol{\bar{x}_{i.}}$ *is the time average of ATL measure for individual i,*

$\bar{\boldsymbol{w}}_{\boldsymbol{i}.}$ *are the time average of problem behaviors for individual i,*

$\bar{\boldsymbol{n}}_{\boldsymbol{i}.}$ *are the time average of time-varying covariates for individual i,*

ψ *is the intercept and* a_i *is the error term from* c_i *for individual i,*

And $v_{it} = u_{it} + a_i$ *is the error term, where* u_{it} *is the error term from the unobserved effects model.*

Note. Variables and parameters are in boldface indicate vectors.

Clustered robust standard errors were obtained to correct for potential heteroskedasticity and correlation in the residuals caused by clustered structure and making fully robust inference. The analysis also included time dummies $\boldsymbol{D_t}$ to account for aggregate changes over time (the reference group was spring kindergarten). Failure to control for time effects can induce serial correlation in the residual u_{it} (Wooldridge, 2010). The interaction between demeaned ATL and SES was added into the model as time-varying covariate to control the potential influence. Feasible generalized least squares (feasible GLS) estimation was conducted to estimate the model. Specially, this study was interested in following parameters: the interaction effects of ATL and problem behaviors (moderation effects of problem behaviors) ($\boldsymbol{\lambda}$**)**, the main effects of ATL and problem behaviors (β) and ($\boldsymbol{\gamma}$) if there were no interaction effects, and the effects of observed time-constant variables on the achievement ($\boldsymbol{\eta}$).

Results

Complete Data Analysis

The descriptive statistics and correlation coefficients of the variables in the study are summarized in Appendix A Table A.2 and Table A.3. The results from the CRE estimation using complete data are presented in Table 1.1, Table 1.2 and Table 1.3. The regression coefficients of interactions between ATL and problem behaviors were non-significant for all three subjects (Table 1.1), which indicated that there were no moderation effects of problem behaviors in the relation between ATL and academic performance. The main effect of ATL was statistically significant for reading ($\hat{\beta}$ = 1.394, p <.05), mathematics ($\hat{\beta}$ = 0.848, p <.05), and science achievement ($\hat{\beta}$ = 0.431, p <.05), when controlling for problem behaviors and other covariates. However, two types of problem behaviors - externalizing problem behaviors (EPB) and internalizing problem behaviors (IPB) did not show significant effects on academic achievement when controlling for learning-related behaviors and other covariates in the model.

Table 1.1 Interactions and main effects in complete data analysis

	Reading	Mathematics	Science
ATL	1.394^{*} (0.116)	0.848^{*} (0.109)	0.431^{*} (0.082)
EPB	-0.162 (0.131)	-0.075 (0.130)	-0.034 (0.099)
IPB	0.135 (0.118)	-0.012 (0.112)	-0.082 (0.087)
ATL×EPB	-0.458 (0.289)	-0.199 (0.259)	0.382 (0.201)
ATL×IPB	0.093 (0.281)	0.173 (0.264)	-0.315 (0.202)

Note. p^{*} *<.05. Clustered robust standard errors are shown in parentheses. Sample size = 5735.*

The CRE model allows the estimations of time-constant variables while getting the same estimations of time-varying variables from the fixed effects estimation. The longitudinal study results indicated that student SES, gender, English learner status, and race significantly impacted academic achievement (see Table 1.2).

Table 1.2 Coefficients of time-constant covariates in complete data analysis

	Reading	Mathematics	Science
SES	3.574* (0.210)	3.297* (0.219)	2.511* (0.158)
Female	-0.177* (0.274)	-6.918* (0.289)	-3.017* (0.208)
Non-English at home	-3.446* (0.463)	-2.047* (0.484)	-3.153* (0.358)
Age	-0.040 (0.030)	0.045 (0.033)	0.055* (0.024)
Race/Ethnicity: Reference group – White students			
Black	-2.077* (0.510)	-8.172* (0.553)	-5.206* (0.393)
Hispanic	-0.636 (0.422)	-3.244* (0.439)	-2.486* (0.323)
Asian	2.133* (0.627)	1.732* (0.650)	-0.434 (0.483)
Other	1.029 (0.550)	-0.671 (0.617)	-0.198 (0.439)

Note. $p^{*} < .05$. *Clustered robust standard errors are shown in parentheses. Sample size = 5735.*

In detail, the students with higher SES got higher scores in reading, mathematics, and science from kindergarten to fifth grade. Female students had lower average scores compared with their male peers, especially in mathematics. English learners had lower academic achievement from kindergarten to fifth grade. The reference group of race and ethnicity in the model is white students. Compared with white students, black and American African students got lower scores in three subjects. Hispanic students had a similar average reading score but lower math and science score. Asian and Asian American students got higher achievement scores in reading and mathematics, but not in science. Students in other races had no significant difference in reading, mathematics and science achievement. Age only shows a significant effect on science performance.

Other covariates in the CRE model were time-varying variables which were controlled in the model. Table 1.3 summarized the estimation results. The results found that time-varying covariates from teacher and school's characteristics (i.e., teacher experience, school enrolment and school SES) did not influence achievement significantly (Table 1.3). Also, students who changed school at each grade and the interaction between ATL and student SES did not significantly impact academic achievement from kindergarten to fifth grade. Time dummies in the model were significant, which indicated that grade effects exist in the model. Controlling the grade effects (time effects) is necessary in the CRE model.

Table 1.3 Coefficients of time-varying covariates in complete data analysis

	Reading	Mathematics	Science
Teacher Experience	-0.002 (0.005)	0.003 (0.004)	0.003 (0.004)
School Enrollment	-0.001 (0.001)	-0.001 (<0.001)	<0.001 (<0.001)
School SES	<-0.001 (0.005)	-0.005 (0.005)	-0.004 (0.003)
Change School	-0.226 (0.217)	0.379 (0.209)	0.266 (0.162)
ATL×SES	-0.081 (0.137)	0.011 (0.126)	-0.014 0.095
Time Dummies: Reference group - Kindergarten Spring			
1^{st} Spring	26.896^* (0.151)	23.118^* (0.117)	9.527^* (0.087)
2^{nd} Spring	44.223^* (0.166)	40.858^* (0.146)	19.199^* (0.105)
3^{rd} Spring	52.472^* (0.172)	54.454^* (0.156)	26.837^* (0.111)
4^{th} Spring	60.593^* (0.165)	63.110^* (0.157)	33.443^* (0.114)
5^{th} Spring	67.692^* (0.178)	70.149^* (0.157)	40.233^* (0.126)

Note. $p^* < .05$. *Clustered robust standard errors are shown in parentheses. Sample size = 5735.*

Bootstrap and Multiple Imputation

Overall, 10702 students were included in the bootstrap and multiple imputation analysis. Those students had their achievement outcomes in reading, mathematics and science in six time periods, but some independent variables (time-varying and time-constant variables) were missing at some time periods. The missing rate of the main predictors (ATL and problem behaviors) is about 10%. The regression coefficients shown in the following tables were the averages from 1000 times bootstrap with multiple imputation and the standard errors came from the bootstrap inference. The CRE model in the section was same as the model in the complete case analysis. The coefficients of the interested variables and time-constant variables from bootstrap with multiple imputation are discussed in the section.

Table 1.4 presents the interaction effects and main effects from bootstrap with multiple imputation. The results for reading and mathematics achievement (coefficient directions and significant levels) were similar to the results from the complete data analysis.

Table 1.4 Interactions and main effects in bootstrap and multiple imputation

	Reading	Mathematics	Science
ATL	1.093* (0.083)	0.736* (0.079)	0.283* (0.059)
EPB	-0.171 (0.096)	-0.155 (0.090)	-0.185* (0.074)
IPB	-0.057 (0.080)	-0.060 (0.081)	-0.057 (0.061)
ATL×EPB	-0.062 (0.202)	-0.036 (0.176)	0.283 (0.154)
ATL×IPB	-0.217 (0.200)	0.046 (0.182)	-0.283* (0.141)

Note. p^* <.05. *Bootstrap standard errors are shown in parentheses. Sample size = 10702.*

The coefficients of problem behaviors were larger than the model using complete data, but there were no interaction effects between ATL and problem behaviors. The coefficients of ATL in the bootstrap sample were smaller than the coefficients in the previous analysis. However, the main effects of ATL were still positive and significant on reading ($\hat{\beta}$ = 1.093, *p*

<0.5) and mathematics ($\hat{\beta}$ = 0.736, p <.05) when controlling for problem behaviors and other covariates in the CRE model. The bootstrap standard errors were smaller than the clustered robust standard errors in complete data analysis. The 95% confidence intervals (95% CI) of the bootstrap coefficients (Table 1.5) showed the range of possible values of the regression coefficients from bootstrap and multiple imputation. The results were consistent with the previous findings.

The results for science achievement were slightly different from the previous results. The interaction between ATL and internalizing problem behaviors was significant ($\hat{\beta}$ = -0.283, p <.05). It indicated that the effect of ATL on science achievement decreased among the students with higher internalizing problem behaviors. The main effect of externalizing problem behaviors was significant on science ($\hat{\beta}$ = -0.185, p <.05), which demonstrated students with higher externalizing problem behaviors would have lower science scores. However, when we look at the 95% confidence intervals from Table 1.5, the upper bound of confidence interval of the interaction term (ATL×IPB) and the main effect (EPB) were very close to zero. Thus, it might suggest the effect of interaction between ATL and internalizing problem behaviors and the main effect of externalizing problem behaviors might not be considerable. When we go back to look at the results in the complete data analysis, these two coefficients were close to the range of 95% confidence interval in the simulation analysis, it demonstrated that the differences are due to the data processing methods. However, the differences were not influential.

Table 1.5 The 95% CI of interactions and main effects in bootstrap and multiple imputation

	Reading	Mathematics	Science
ATL	[0.930, 1.256]	[0.582, 0.898]	[0.166, 0.399]
EPB	[-0.360, 0.018]	[-0.320, 0.022]	[-0.330, -0.040]
IPB	[-0.213, 0.099]	[-0.219, 0.099]	[-0.178, 0.063]
ATL×EPB	[-0.459, 0.334]	[-0.382, 0.310]	[-0.020, 0.585]
ATL×IPB	[-0.609, 0.176]	[-0.310, 0.401]	[-0.559, -0.007]

Note. Sample size = 10702.

Figure 1.1 to Figure 1.3 display the distributions of ATL coefficients from bootstrap with multiple imputation on different academic achievement (reading, mathematics and science). The distributions showed the range of bootstrap ATL coefficients. It also suggests that the data-based simulation resampling method works on the problem and the coefficients are normally distributed because of a large number of repeated times (1000 times).

Figure 1.1 The distribution of ATL coefficients from bootstrap for reading

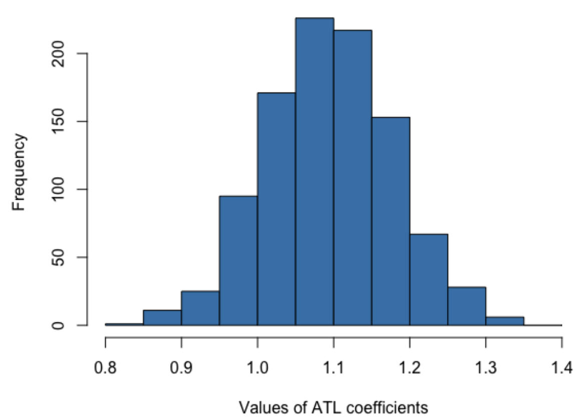


Figure 1.2 The distribution of ATL coefficients from bootstrap for mathematics

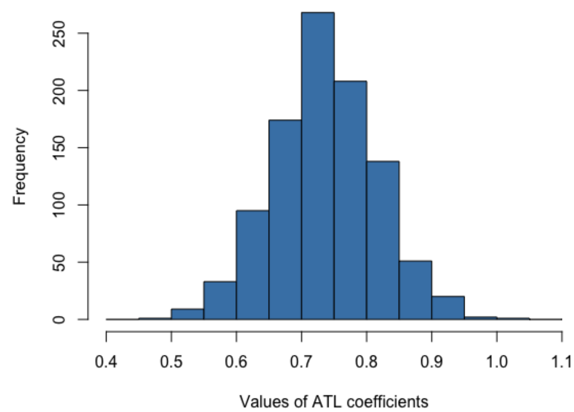
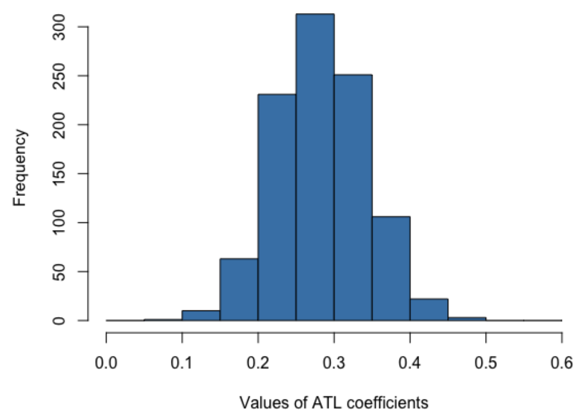


Figure 1.3 The distribution of ATL coefficients from bootstrap for science



The results from Table 1.6 indicates the coefficients and standard errors of time-constant variables in bootstrap with multiple imputation analysis. The results (coefficient directions and significant levels) were very similar to the results from the complete data analysis. The effect of age was significant for all three subjects, which differed from the results in the complete case analysis. However, compared with other time-constant covariates, the size of age effect was not large.

Table 1.6 Coefficients of time-constant covariates in bootstrap and multiple imputation

	Reading	Mathematics	Science
SES	3.646* (0.168)	3.920* (0.178)	2.545* (0.121)
Female	-1.393* (0.195)	-6.535* (0.217)	-3.000* (0.160)
Non-English at home	-2.045* (0.308)	-1.131* (0.321)	-2.787* (0.237)
Age	0.072* (0.023)	0.162* (0.025)	0.150* (0.018)
Race/Ethnicity: Reference group – White students			
Black	-1.910* (0.384)	-7.200* (0.407)	-5.012* (0.266)
Hispanic	-1.122* (0.305)	-3.280* (0.319)	-2.502* (0.154)
Asian	1.667* (0.373)	2.520* (0.409)	-0.658* (0.279)
Other	0.601 (0.464)	-0.860 (0.479)	-0.478 (0.337)

Note. $p^* < .05$. *Bootstrap standard errors are shown in parentheses. Sample size = 10702.*

The bootstrap standard errors were smaller than the clustered robust standard errors in complete data analysis. The 95% confidence intervals from Table 1.7 also indicated the similar results. The findings confirmed that demographic variables are strongly influence the academic trajectories of reading, mathematics and science in early grades in recent years.

Table 1.7 The 95% CI of time-constant covariates coefficients in bootstrap and multiple imputation

	Reading	Mathematics	Science
SES	[3.318, 3.975]	[3.043, 3.740]	[2.308, 2.782]
Female	[-1.775, -1.010]	[-6.961, -6.110]	[-4.391, -2.683]
Non-English at home	[-2.649, -1.441]	[-1.759, -5.027]	[-3.252, -2.322]
Age	[0.026, 0.118]	[0.133, 0.211]	[0.080, 0.150]
Race/Ethnicity: Reference group – White students			
Black	[-2.663, -1.157]	[-8.000, -6.403]	[-5.534, -4.490]
Hispanic	[-1.719, -0.524]	[-3.906, -2.655]	[-2.934, -2.071]
Asian	[0.935, 2.398]	[1.717, 3.322]	[-1.204, -0.112]
Other	[-0.309, 1.511]	[-1.800, 0.079]	[-1.137, 0.182]

Note. Sample size = 10702.

Discussion

The empirical study in Chapter 1 investigated the moderation effect of problem behaviors on the relationship between ATL and achievement using a recent longitudinal dataset in education. The CRE model was applied in the study to control the omitted bias issue better. At the same time, the model could provide the estimations of the effects of critical time-constant variables (e.g., demographic variables) on the outcomes. The study conducted two parts of analyses, complete data analysis and bootstrap with multiple imputation analysis. The second analysis aimed to deal with the missing data issue and showed the possibility of the strategy in panel (longitudinal) data analysis.

The results from complete data analysis and bootstrap with multiple imputation indicated no significant interactions between problem behaviors (externalizing and internalizing) and ATL on reading and mathematics achievement from kindergarten to fifth grade. In other words, the moderation effects of problem behaviors non-significantly impacted the relationship between ATL and academic achievement. It indicated that students with different degrees of problem behaviors had a similar relationship between their learning-related behaviors and cognitive testing scores. However, the main effects of ATL were significant when controlling for problem behaviors, which suggested that ATL was strongly associated with academic achievement in early grades. The results were consistent with the previous findings that ATL was an important indicator for academic trajectories in childhood (Li-Grining et al., 2010; McClelland et al., 2006). The findings also showed that the effect of ATL on reading achievement was more significant than the effect on mathematics. However, the main effects of problem behaviors on achievement were non-significant when controlling for ATL. The finding was also consistent with the result from McWayne's study (2004), which showed that behavioral problems did not influence the academic success significantly when controlling ATL among students in preschool. The present study extended the results to elementary grades and provided a robust evidence from longitudinal perspective. The results from two statistical analyses on science achievement showed some differences. One possibility is that the internalizing problem behaviors affected the relationship between ATL and science performance. However, the moderation effect was around a significant level. It might be more sensitive to the data size in this case. Therefore, the effect was detected only by the complete data analysis.

Based on the results from a nationwide large-scale educational data with a group measure of ATL, the findings showed more convincing evidence that children should have some instruction about "how to learn" during their school years to help them achieve higher cognitive performance. Considering "ATL frame the child's entire being and are at the core of social/emotional and cognitive interactions (Kagan et al., 1995)", the empirical results suggest that ATL is worthwhile to gain more attention from educators and policymakers.

From a practical perspective, the findings of this study imply that interventions and training are important to help students build learning-related skills in early grade. Indeed, teachers and parents could play a crucial role in improving ATL. Previous studies have found that learning-related skills could be improved in daily learning activities in the classroom or at home. For example, tutoring inattentive students helped them perform better in reading (Rabiner et al., 2004). Moreover, students could be trained to develop self-regulation skills during homework activities (Ramdass & Zimmerman, 2011). Students whose parents participated more in a learning-related behavioral intervention got better outcomes to a greater extent (McCormick et al., 2016). An eight-week class-based intervention on self-regulation was helpful for children to enhance school readiness and improve academic achievement in preschool, especially for English language learners (Schmitt et al., 2015). A famous intervention is called Tools of Mind, which was designed to foster children's regulation skills and attention (Bodrova & Leong, 1996; Bodrova & Leong, 2019). Thus, it could improve children's skills of social competency and ATL. The Tools of Mind curriculum applies in class with regular teachers. Teacher guide 40 small activities, such as self-regulatory private speech, dramatic play, and provide dynamic instructional and emotional support depending on children's accomplishment (Diamond et al., 2007). As a result, both teachers and parents are presented with great opportunities to facilitate learning approaches during routine learning activities either in school or at home. Although focusing on effective teaching and instruction are key enablers of learning, the results suggest that helping students build great learning-related behaviors continuously in early grades is important as well. Hence, it may be beneficial to encourage educators and parents to provide appropriate training to students in early grades to improve their learning-related skills.

The present study applied a CRE model, which could get the same estimations of time-varying variables from a fixed effects estimation model. Meanwhile, the important time-constant variables could be evaluated in the same model. Student SES, gender, English learner status and race/ethnicity were included in the model to investigate the effects on academic achievement from the recent longitudinal large-scale dataset. The results presented that most demographic variables significantly influenced academic performance. Specifically, student SES showed a substantial effect on reading, mathematics and science achievement. Female students showed a lower score in mathematics than male students significantly. English learners had lower academic performance in three subjects. Also, the students with different races/ethnicities performed differently in three subjects in early grades. The results suggested that those differences in demographic variables related to students' academic and cognitive performance in the current education systems. Educators and policymakers need to keep reforming education to close the gaps.

The study also used a strategy to deal with missing data from the dataset. Multiple imputation is widely used in the research with missing data. The bootstrap is a statistical inference method based on resampling from the data. Each coefficient after bootstrap has a distribution and the bootstrap standard error is computed to support a robust inference. The combination of these two methods helps to get a robust result when dealing with the missing data. The comparison between the coefficients from two analyses (complete data analysis and bootstrap with multiple imputation) in the chapter showed the similar results for the main variables, including the directions and significant levels of the regression coefficients. The bootstrap standard errors were smaller than the robust clustered standard errors. It is possible because when applying multiple imputation to deal with missing data issues, the total sample size increases. Overall, the results suggested that the complete data analysis could reflect the existing effects on academic growth. Also, it demonstrated that the bootstrap with multiple imputation works for the panel data analysis.

Although the coefficient directions and statistical inferences were very similar in the two approaches, the values of coefficients from the two approaches possibly had some differences. For instance, the coefficients of interactions and main predictors were smaller in the second approach with more minor standard errors. Future work could include additional tests to test the value differences between the same coefficients from complete case analysis and bootstrap with multiple imputation. Advanced tests might be involved due to the dependence of the coefficients. Additionally, the Hausman test (Hausman, 1978) could be added to compare the differences statistically between regression coefficients in different methods (fixed-effects vs. random-effects), thus determining the best approach for the data.

Future work could also consider more complex structures based on the CRE model. For instance, it is possible to model fixed effects as having varying effects over time, such as testing the time-varying effects of the time-constant covariates in the model. Potential mediators or moderators (e.g., psychological functioning) might be taken into account in the relationship between ATL and academic achievement in childhood.

CHAPTER 2 A META-ANALYTIC REVIEW ON THE RELATIONS BETWEEN APPROACHES TO LEARNING AND ACADEMIC ACHIEVEMENT IN CHILDHOOD FROM QUASI-EXPERIMENTAL EVIDENCE

Introduction

The previous chapter focused on a longitudinal study of ATL and achievement considering moderation effects of problem behaviors. Although previous evidence has demonstrated that the components of ATL had strong associations with academic achievement, DiPerna and Elliott (2002) suggested building a more comprehensive model to understand contributions of the combinations of enablers (including learning approaches). The present chapter extended the external validity of the studies. A systematic review with meta-analysis was conducted to get a general understanding of the relationship between ATL and academic achievement in childhood from quasi-experimental evidence. Meta-analysis is widely used in psychology, social science, and medicine. It is a quantitative method to summarize the results of several empirical research studies from similar topics (Hedges, 1992). It refers to statistical modeling in systematic reviews. Meta-analysis offers a rigorous methodology for quantitative research synthesis, follows specific guidelines/criteria, and has structured processes. Thus, it has high external validity and greater statistical power from measurement perspectives, and it is considered an evidence-based resource.

One prior meta-analysis was conducted to detect the effects of learning-related skills interventions on student learning in the late 90s. Hattie and co-researchers (1996) found a mean weighted effect size of learning-related skills on achievement was 0.45 with a standard error of 0.03. The effect size indicates a medium effect size. Moderation effects of age and academic ability were found in their study. In detail, the interventions conducted in primary schools showed the strongest effect size, and students with medium ability showed the strongest effect size. However, that study focused on intervention designs. Only a small proportion of studies in the meta-analysis had a similar definition of ATL. Also, the meta-analysis was published nearly 25 years ago, new and recent evidence is not available. Additionally, some previous meta-analyses mainly focused on one specific component of ATL and how to improve it, but they did not test the effects on academic achievement. For instance, Dignath et al. (2008) examined the effects of students learn self-regulated learning strategies in elementary school on several self-regulation training programs. Therefore, to fill the gap, the present meta-analysis using recent evidence would provide a general view of ATL components' associations and achievement in early grades among quasi-experimental designs.

Literature Review

As mentioned in Chapter 1, ATL faded in educational researchers' sight at the end of the last century. ATL was considered as the least research domain for school readiness (NEGP, 1991). However, ATL as a general domain related to learning might be the most critical indicator for school readiness because the components of ATL might serve as "causal protective resilience factors during the transition to school" (George & Greenfield, 2005, p.70). The U.S. Department of Health and Human Services also suggests ATL as a key domain contributing to school success directly (U.S. Department of Health and Human Services, 2019). To emphasize that ATL is separated from social-emotional learning as an independent school readiness domain by the framework. The following paragraphs in the section review the common measurement scales of

ATL components and the prior findings of the relationship of ATL and academic achievement in early grades from different quasi-experimental research designs.

How to Measure

Prior studies mainly used two tools to measure children's ATL or learning-related behaviors. Atkins-Burnett developed a rating scale of ATL for ECLS in the 1990s. The rating scale has been used in two rounds of ECLS assessment (ECLS-K:1998 and ECLS-K:2011) for measuring students' learning approaches in early grades. Studies used ECLS datasets (e.g., Li-Grining et al., 2010; Tach & Farkas, 2006; Robinson & Mueller, 2014) usually choose the ATL composite score as a measure of ATL. The composite score was computed by the ECLS research team considering the missing rate of the items. The ATL instruments include seven components related to behaviors, inclinations, and dispositions during learning activities. The ECLS datasets provide both teacher rating and parent rating score of ATL. Some studies (e.g., Razza et al., 2015) extracted and adjusted the ECLS scale of ATL and used it to measure ATL in their own studies.

On the other hand, ATL was measured by learning behaviors scales in previous studies. Preschool Learning Behaviors Scale (PLBS) was developed by McDermott et al. (2000) to assess 3 to 5-year-old preschooler's learning-related behaviors. Three dimensions, including Competence Motivation, Attention/Persistence, and Attitude Toward Learning, are measured by 29 items. Further, McDermott and co-researchers tested its validation and evidenced that the scale provided a structured and robust measure of learning-related behaviors (McDermott et al., 2002; McDermott et al., 2012). Learning Behaviors Scale (LBS) is similar to PLBS, but it was developed for older children in kindergarten, elementary and secondary school (McDermott, 1999). Compared with PLBS, it has one more dimension, which is called Strategy/Flexibility. Both scales are teacher rating scales. Other researchers (e.g., McWayne et al., 2004; Durbrow et al., 2001; Rikoon et al., 2012) applied these scales to their studies as a measurement tool to analyze ATL.

Besides these two popular scales, some studies chose other scales or methods to measure ATL. Stipek et al. (2010) extracted four items from the Teacher Rating Scale of School Adjustment (TRSSA; Birch & Ladd, 1997) to measure students' learning-related behaviors in elementary school. McClelland et al. (2006) used a subscale from the Cooper-Farran Behavioral Rating Scales (CFBRS; Cooper & Farran, 1991) as a measure of children's learning-related skills. Williams et al. (2016) extracted ATL-related items from the Social Skills Rating Scale (SSRS; Gresham & Elliott, 1990) and defined the components under the attentional/cognitive regulation dimension. George and Greenfield (2005) designed a structured problem-solving flexibility task to reflect ATL levels. They have demonstrated that the task score was significantly correlated to the teacher rating ATL score.

Single Timepoint Evidence

Previous findings indicated a significant association between ATL and academic achievement from single timepoint analyses using diverse samples. Bustamante and Hindman (2019) found ATL could directly influence preschooler's academic readiness when testing the relationship between classroom quality and academic readiness using Family and Child Experiences Survey data. George and Greenfield's study (2005) suggested the strong influence of ATL on concurrent achievement in kindergarten. Children from the Fragile Families and Child Wellbeing Study showed their ATL significantly impacted reading and mathematics achievement at age 5 (Razza et al., 2015). Several studies selected subsamples from the ECLS-K dataset in different grades. One study showed that ATL and mathematic outcomes were significantly correlated in kindergarten when controlling class-level covariates into the model (Robinson & Mueller, 2014). Bumgarner et al. (2013) showed a positive relationship between ATL and mathematics achievement among Hispanic immigrant children (known as English language learners) in kindergarten, first grade, and third grade.

Longitudinal Evidence

The potential positive relationship between ATL and academic achievement in childhood was found from longitudinal evidence. Li-Grining and co-researchers (2010) used the ECLS-K:1998 dataset to investigate the impact of early ATL on academic performance. The results from the large-scale assessment demonstrated that ATL at kindergarten entry was significantly associated with reading and mathematics achievement trajectories through fifth grade. McClelland and colleagues (2006) showed a similar result using a different sample that early learning-related behaviors at kindergarten strongly impact reading and mathematics growth through elementary school years. Williams et al. (2016) found that ATL at 6-7 years of age predicted later mathematics achievement at 8-9 years of age. Other results indicated that prior learning-related behaviors in early elementary school years could predict later literacy performance among students from low-income families (Stipek et al., 2010). Also, research showed similar results of the relationship of early ATL and later academic achievement when considering different subgroups from the ECLS-K dataset (Tach & Farka, 2006; Mattew et al., 2010).

Potential Moderators in the Relation

Some moderators were introduced and tested in the relationship between ATL and academic achievement in early grades. Robinson (2013) provided results that poor or low-income students could moderate the effect of behavioral engagement on mathematics gains. The finding suggested that it could be beneficial for poor students with high behavioral engagement on achievement. Second, gender moderated the relation between ATL and academic performance. Li-Grining et al. (2010) showed that ATL at kindergarten was more protective for female students' mathematics growth and male students' reading growth in elementary school years. Mattews et al. (2010) indicated a significant interaction effect among race, gender and ATL on reading achievement. In addition, academic competence at early ages could be a potential moderator from previous evidence. The studies found that ATL would benefit more on later academic achievement for students with low academic skills in early grades (Razza et al., 2015; Li-Grining et al., 2010). Other possible moderators were shown some evidence from previous studies. For instance, the moderation effect of English proficiency existed in kindergarten and third grade (Bumgarner et al., 2013). Additionally, class and school level moderators possibly existed, such as the frequency of reading activities in class and school enrollment (Musu-Gillette et al., 2015).

ATL as A Mediator or Moderator

When testing potential predictors for academic achievement in early grades, ATL was used as a mediator or moderator in research. ATL was investigated as a mediator for the relationship between psychological functioning and academic achievement in childhood. For instance, Sánchez-Pérez and colleagues (2018) found the mediation effect of ATL on effortful control and reading/mathematics performance in elementary school. ATL was indicated as a mediator in the relationship between cognitive flexibility and academic school readiness for Headstart children in preschool (Vitiello et al., 2011). Moreover, ATL mediated the relation between children's executive function skills and concurrent and later academic achievement (Nesbitt et al., 2015; Sasser et al., 2015).

Other prior research focused on testing the associations of parenting characteristics and student academic performance. ATL was found as an important mediator of these associations. Smith-Adcock et al. (2019) targeted students with low socioeconomic scores and showed that ATL has a mediation effect between parenting stress and reading achievement in kindergarten. ATL could be a significant mediator for divorce and academic achievement in elementary grades (Anthony et al., 2014). Additionally, studies showed that school-level involvement was indirectly associated with achievement through ATL (Anthony and Ogg, 2019; Smith-Adcock et al., 2019). ATL moderated the relationship between classroom quality and writing/spelling skills among Head Start children (Meng, 2015).

Present Study

The aim of this study was to fill in this gap in the literature, to conduct a systematic review with meta-analyses to detect an average effect of ATL (learning-related behaviors) on reading and mathematics achievement in childhood from different quasi-experimental study designs (i.e., single timepoint analysis, longitudinal analysis). Specifically, the study addressed the following research questions:

(1) Is there a significant relationship between ATL and achievement in childhood from quasi-experimental designs?

(2) How large is the average effect of ATL on student achievement from quasi-experimental designs?

(3) What kind of variables could moderate the effect on achievement?

Therefore, the present study conducted a systematic review employing meta-analytic methodology to combine and summarize the quasi-experimental results of empirical research studies about the relation of ATL and achievement approximately from 2000 through 2020. Detailly, four meta-analysis conditions were conducted in the study: single timepoint results for reading achievement, single timepoint results for mathematics achievement; longitudinal results for reading achievement, and longitudinal results for mathematics achievement. The study extended the theory and understanding of the relations between ATL and achievement in recent years by using the meta-analysis method. The study could get more clear results because of including both one timepoint and longitudinal results. For practical significance, results from the present study could help researchers, educators, and policymakers make decisions to use proper ATL educational programs under a current education environment.

Methods

Literature Search

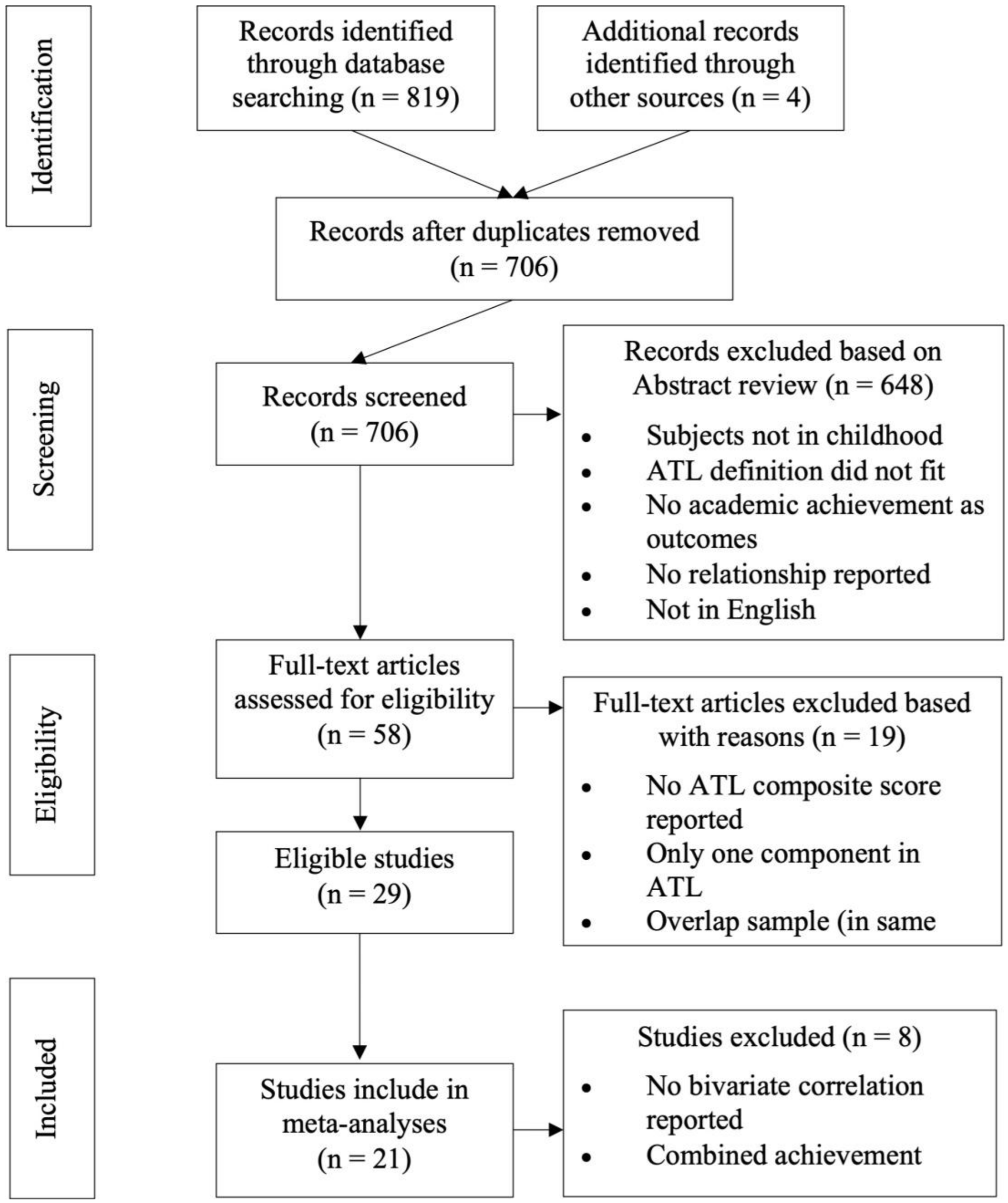
This study aimed to conduct a meta-analysis about the relationship between ATL and student achievement (reading and mathematics) in childhood in the recent 20 years (2000-2020) from quasi-experimental designs. The meta-analysis used quasi-experimental evidence because few interventions directly focused on combined ATL components and the results are hard to classify from designs with other components (e.g., components from social-emotional learning, problem behaviors, social competence, or class management). A computer search of potential databases, including Web of Science, ERIC and PsycINFO, was conducted to identify the relevant literature using keywords, "approaches to learning; learning behaviors; achievement" or "learning-related behaviors; achievement". The possible year range was from 2000 till 2020. The age group focuses on childhood (preschool to elementary school). The initial literature search yielded 819 studies with over ten dissertations. Additional four possible studies came from references of relevant papers. After getting an initial study pool, 113 non-relevant and duplicated studies were excluded from the pool.

Study Selection Criteria

A detailed protocol was created in the study to define explicitly the criteria for including and excluding studies and to create a final sample of studies eventually. After getting the initial study pool, a screening phase selected the studies by reviewing abstracts. The studies were excluded in the screening phase because 1) they were not written in English; 2) the definitions of ATL or learning-related behaviors did not fit in the current analysis; 3) there was no appropriate reading and mathematics achievement score reported in the study; 4) participants were not in childhood; 5) no relationship of ATL and achievement was reported in the study. Fifty-eight full-text articles were eligible after screening. The eligibility phase excluded several studies after full-text reading. The first reason is that no ATL composite score was used to test the relationships (using components separately into the analyses). Second, there was only one component representing ATL in the study.

The present study decided to extract bivariate correlations as effect sizes in the final phase because quasi-experimental designs did not report mean differences as intervention studies. The Pearson correlation is one of the most common and important effect sizes used in meta-analysis (Rosenthal, 1994; Rosenthal, 1995). The studies without the selected statistics were excluded from the final sample. The studies which reported the standardized regression coefficients or used a combined achievement test only were included in the study report table (see Appendix B Table B.1). However, they were not included in the meta-analysis because different and complex model designs make the coefficients incomparable. There were several extra selection criteria at the final stage. If the study provided a range of correlation coefficients, the midpoint was used to represent the correlation coefficient of the study (only one study in the pool). If two studies had very similar sample (same survey and participants at the same grade), the effect size with smaller sample size was excluded in the study. If a study had more than one independent sample, the effect sizes were included in the study. Overall, 21 studies were included in the final sample for meta-analysis. Figure 2.1 is a flow chart to show the detailed procedures of choosing final meta-analysis samples in the study from searching, initial screening, to eligibility and final selection.

Figure 2.1 A flowchart of searching and screening results



Statistical Analysis

Fixed-effects and random-effects model in meta-analysis (Hedges & Olkin, 2014; Borenstein et al., 2007) were applied in the present study. The fixed-effects model could be treated as “a linear weighted regression” and assumes all studies estimate the same true effect size. The method generalizes studies in the sample. Compared with the fixed-effects model, the random-effects model assumes each study is estimating a unique effect. The random sample is from a larger population. It provides a more general statement and gets inference from the sample. In other words, the random-effects model generalizes to a larger population of studies. All analyses applied Fisher’s Z transformation to eliminate the potential bias from correlation coefficients. The transformation provides a correction for a skewed sampling distribution of correlations (Fisher, 1921).

Fixed-Effects Model

The observed effect size in study *i* equals to a sum of a true (population) effect size and within-study error from the fixed-effects model, which is shown as

$$T_i = \mu + \varepsilon_i. \tag{2.1}$$

The model is an intercept only linear regression model. The variances of error term are assumed known. Thus, by using the weighted linear regression estimation method, the average weight effect size $\bar{T}.$ from *k* studies could be calculated by

$$\bar{T}. = \frac{\sum_{i=1}^{k} w_i T_i}{\sum_{i=1}^{k} w_i}, \tag{2.2}$$

where w_i is the inverse of the within-study variance (v_i) for study *i*. The standard error of the average weighted effect size in the fixed-effects model is shown as

$$SE(\bar{T}.) = \sqrt{v.} = \frac{1}{\sqrt{\sum_{i=1}^{k} w_i}}. \quad (2.3)$$

Random-Effects Model

The observed effect size in study *i* equals to a sum of a true effect, a between study error and a within-study error in the random-effects model, which is shown as

$$T_i = \mu_i + \zeta_i + \varepsilon_i. \quad (2.4)$$

The model is an intercept only linear regression model. The differences between Equation (2.1) and Equation (2.4) are that each study has its own true effect (μ_i) and the between-study error (ζ_i) is introduced into the model. Thus, the random effects model considers heterogeneity between studies. The weighted average effect size ($\bar{T}.^*$) from *k* studies is calculated by the new weights (w_i^*), which include two parts of variance: within-study variance (v_i) and between-study variance (τ^2). The equation is represented as:

$$\bar{T}.^* = \frac{\sum_{i=1}^{k} w_i^* T_i}{\sum_{i=1}^{k} w_i^*}, \quad (2.5)$$

where $w_i^* = \frac{1}{v_i^*} = \frac{1}{v_i + \tau^2}$.

The new variance (v_i^*) is the sum of the within-study variance for study *i* and the between-study variance. The between-study variance is estimated by restricted maximum likelihood estimation method (REML) in the study. And the standard error can be calculated by

$$SE(\bar{T}.^*) = \sqrt{v.^*} = \frac{1}{\sqrt{\sum_{i=1}^{k} w_i^*}}. \quad (2.6)$$

Heterogeneity Tests

To determine which meta-regression model fits the data better, a heterogeneity test should be conducted. The null hypothesis of the test is that all population effect sizes are same (the homogeneity of population effects sizes). The Q statistics could be calculated to test the hypothesis. The Cochran's Q test follows the results of the fixed-effects model, which is shown as:

$$Q = \sum_{i=1}^{k} w_i (T_i - \bar{T}_{.})^2, \tag{2.7}$$

where $\bar{T}_{.}$ is the weighed effects size in Equation (2.2). The Q statistics follows a chi-square distribution with $k - 1$ degrees of freedom.

The I^2 statistics represents the proportion of total variation due to heterogeneity (Higgins & Thompson, 2002). The statistics could be calculated using Q statistics to quantify inconsistency across studies. The larger value of I^2 indicates a larger amount of heterogeneity across the studies. The I^2 index can be computed from

$$I^2 = \frac{(Q - df)}{Q} \times 100\%. \tag{2.8}$$

Moderation Analysis

In additional to computing the weighted effect sizes, the study examines differences in individual studies (i.e., study characteristics) as well. This is called a moderation analysis. The model could be called meta-regression model because predictors (moderators) are in the model. The moderation analysis indicates regressing effect sizes (outcomes) on the study characteristics (moderators) (Hedges & Olkin, 2014). Suppose that each effect parameters are determined by p moderator variables $X_1, X_2 \ldots X_P$.

The fixed effects model with moderation analysis is shown as:

$$T_i = (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}) + \varepsilon_i. \tag{2.9}$$

And the random effects model with moderation analysis is shown as:

$$T_i = (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \zeta_i) + \varepsilon_i. \tag{2.10}$$

The study is interested in estimating parameters (β) from the fixed-effect model if the homogeneity assumption is met or the random-effects model if the homogeneity assumption is violated. Parameters (β) reflect the effects of moderators chosen in the study.

Potential moderators in the present study were the year of the study, grade, socioeconomic status (SES) of students, and publication type. Grade was a categorical variable with three categories (preschool: 0; kindergarten: 1; elementary school:2). Student SES/Minority was coded as a binary variable. The reference group was regular students, and the other group was disadvantaged students (i.e., low income, low SES or minority). The publication type was binary variable which indicated the study was from a peer-reviewed paper or a dissertation. The variable of year was centered to the mean and was treated as a continues variable.

Sensitivity Analysis

Because a few studies from the pool could extract more than one effect size (correlation), the present study needed to decide on how to deal with multiple effect sizes within one study. Thus, a sensitivity check was conducted first. The sensitivity check aimed to determine which analysis approach (univariate meta-analysis or multivariate meta-analysis) would be applied for the final study pool. A sensitivity analysis can acknowledge the dependence issue by applying analyses using all outcomes in each study and using one outcome in each study. If the results from two approaches are similar, it makes sense to drop or combine multiple effect sizes within the study (Becker, 2000). First, the sensitivity check did meta-analysis with the full sample and assuming correlations of the same study are independent to each other. Second, for the studies with multiple correlations, the sensitivity checking randomly kept one correlation as the effect size of the study and did meta-analysis with the subsample. By comparing the results from the two procedures, we could have an understanding about how large the multiple effect sizes within one study influence the final results (weighted average effect sizes). The initial results are shown in Table 2.1 to indicates the meta-analysis results from sensitivity check.

Table 2.1 The results from sensitivity check

		Single timepoint designs		Longitudinal designs	
		Reading	Mathematics	Reading	Mathematics
Fixed-effects model	Full sample	0.403^{*} [0.396, 0.408]	0.327^{*} [0.320, 0.336]	0.400^{*} [0.395, 0.404]	0.390^{*} [0.384, 0.396]
	Subsample	0.420^{*} [0.414, 0.426]	0.333^{*} [0.325, 0.341]	0.385^{*} [0.377, 0.392]	0.377^{*} [0.368, 0.385]
Random-effects model	Full sample	0.338^{*} [0.274, 0.399]	0.309^{*} [0.228, 0.386]	0.374^{*} [0.338, 0.409]	0.357^{*} [0.312, 0.401]
	Subsample	0.365^{*} [0.292, 0.433]	0.342^{*} [0.259, 0.420]	0.346^{*} [0.275, 0.413]	0.328^{*} [0.244, 0.408]
	Full sample N	23	15	20	12
	Subsample N	16	12	8	6

Note. The 95% of confidence intervals are shown in the brackets. $p^{*} < .05$.

All tests of heterogeneity were significant, which indicates that the random-effects model fitted the data better. The results showed the weighted average effect size changed slightly in each condition (different research designs and achievement outcomes). In random-effects models, the weighted effect sizes were in the range of 95% confidence intervals of the weighted effect sizes in full sample analysis. The magnitude of the weighted effect size did not change (e.g., from medium to small effect size). The evidence provides an argument that univariate meta-analysis could be an appropriate design in the present study. Although multivariate approach has been developed fast in recent years, it still has limitations. For instance, compared to univariate meta-analysis, multivariate approach is more complex and harder to understand/interpret; additional assumptions (e.g., multivariate normality) are hard to verify; estimators' statistical properties can only be improved slightly (Jackson et al., 2011). Also, the quasi-experimental designs possibly used a large-scale dataset. The number of participants in each study may be very different. The multivariate analysis possibly hides the true weighted effect size when including effects size from one study with large sample size, especially in the fixed-effects models. Therefore, the present study decided to compute a single effect size in each study and use the univariate meta-analysis approach to yield the final statistical results.

The study firstly applied a fixed-effects model, then conducted a heterogeneity analysis to detect potential significant heterogeneity of effect sizes across studies and figured out if between-study variability should be included in the analysis. If so, the study conducted a random effects model which assumes an effect size is nested within a study. After estimations, the study compared the weights from these two models (fixed and random effects) across all studies in the final sample and determines the contribution of the between-study variance in the weights in the random effects model.

Combining Multiple Correlations

When conducting a univariate meta-analysis, only one effect size should be contained in each study. Multiple effect sizes within the same study needed to be combined into one effect size. The correlation coefficient was treated as the effect size in the present study. The study presented two ways to combine correlation coefficients. A general way to average correlation coefficients from repeated measure was to use Fisher's Z transformation (Silver & Dunlap, 1987). The main procedures include transforming correlations to Fisher's Z score, taking an average Z score, transforming back to a correlation coefficient. This approach demonstrated that the average coefficient was less biased than the untransformed average correlation (Silver & Dunlap, 1987; Strube, 1988). Fisher's Z score transformation for *i*th correlation (r_i) is shown as

$$z_i = \frac{1}{2}\ \ln\left(\frac{1+r_i}{1-r_i}\right). \tag{2.11}$$

Then, we could compute the average z from *k* studies score using

$$\bar{z} = \frac{\sum_{i=1}^{k}(n_i-3)z_i}{\sum_{i=1}^{k} n_i - 3k}. \tag{2.12}$$

After getting the average z score, we use Fisher's Z transformation to transfer back to the correlation coefficient. The transformation from z score to correlation is

$$\bar{r}' = \frac{e^{2\bar{z}}-1}{e^{2\bar{z}}+1}. \tag{2.13}$$

Another approach under meta-analysis research settings was to compute an approximately unbiased minimum-variance estimator (Olkin & Pratt, 1958). The estimator was less biased than the previous one (Viana, 1982; Alexander, 1990). The equation is shown as

$$\bar{r}^* = \frac{\sum_{i=1}^{k}(n_i-1)}{\sum_{i=1}^{k} n_i - k}\{r_i + \left[\frac{r_i(1-r_i^2)}{2(n_i-3)}\right]\}. \tag{2.14}$$

For the studies with multiple correlation coefficients in the present study, the results using the above approaches were very similar. All differences were about or smaller than 0.001. The study rounded the combined effect size to two decimals. Thus, the values of combined effect sizes were same from the two approaches.

Results

Weighted Average Effect Sizes

Four conditions were considered in the study: single timepoint design for reading or mathematics achievement and longitudinal design for reading or mathematics achievement. The results from the fixed-effects and the random-effects model are presented in Table 2.2. All weighted effect sizes were significant than zero in both approaches. The results of heterogeneity tests indicated that there is a large amount of heterogeneity under each condition. Thus, the random-effects models fitted the data better. The results were consistent with the findings in the sensitivity analysis. The weighted effect size was 0.366 in the relationship between ATL and reading achievement from single timepoint designs, which indicated a medium effect size. The weighted effect size was 0.340 in the relationship between ATL and reading achievement from longitudinal designs, which was slightly smaller than the effect size from single timepoint designs. However, it still showed as a medium effect size.

Compared with the results for reading achievement, the effect sizes of the relationship between ATL and mathematics achievement were smaller. Under the condition of single timepoint designs, the weighted effect size was 0.338. And under the condition of longitudinal designs, the weighted effect size was 0.328.

Additionally, the number of studies using longitudinal designs was smaller than the studies for testing the concurrent relationships. And the number of studies for reading was larger than the number of studies for mathematics. Forest plots under four conditions are shown in Appendix B.

Table 2.2 Meta-analysis results

		Single timepoint designs		Longitudinal designs	
		Reading	Mathematics	Reading	Mathematics
Fixed-effects model	ES	0.414*	0.334*	0.377*	0.373*
	95% CI	[0.408, 0.420]	[0.326, 0.342]	[0.369, 0.385]	[0.364, 0.382]
Heterogeneity tests	Q	502.54*	605.85*	329.17*	164.73*
	I^2	97.0%	98.2%	97.9%	97.0%
Random-effects model	ES	0.366*	0.338*	0.340*	0.328*
	95% CI	[0.297, 0.430]	[0.253, 0.418]	[0.272, 0.406]	[0.243, 0.408]
Study N		16	12	8	6
Participant N		69904	45905	44018	32262

Note. $p^* < .05$.

Subgroup Differences

The study conducted two tests for testing subgroup differences separately to show how the weighed effect size influenced by students' SES level or grade level. The subgroup differences tests were under the random-effects models because of the large heterogeneities. Table 2.3 shows that there was no significant difference of the relationship between ATL and academic achievement among the students with or without disadvantages (i.e., low SES, minority). Second, the grade level significantly impacted the relationships in single timepoint designs (reading and mathematics) and longitudinal designs (mathematics). Lower weighted effect size was shown among preschoolers. The studies which focused on kindergarten and elementary school had similar weighed effect sizes.

Table 2.3 Subgroup differences tests results

		Single timepoint designs		Longitudinal designs	
		Reading	Mathematics	Reading	Mathematics
SES	Regular	0.383 [0.329, 0.436]	0.335 [0.251, 0.414]	0.352 [0.233, 0.460]	0.268 [0.107, 0.415]
	Low/Minority	0.338 [0.141, 0.510]	0.350 [0.075, 0.576]	0.321 [0.242, 0.396]	0.378 [0.333, 0.422]
Between group difference		$p = 0.643$	$p = 0.913$	$p = 0.663$	$p = 0.166$
Grade	Preschool	0.194 [0.114, 0.271]	0.238 [0.176, 0.299]	0.289 [0.142, 0.423]	0.170 [0.093, 0.245]
	Kindergarten	0.430 [0.345, 0.508]	0.472 [0.383, 0.553]	0.352 [0.258, 0.438]	0.400 [0.389, 0.411]
	Elementary	0.420 [0.367, 0.471]	0.301 [0.139, 0.447]	0.372 [0.270, 0.466]	0.359 [0.289, 0.424]
Between group difference		$p^* < .05$	$p^* < .05$	$p = 0.625$	$p^* < .05$

Note. 95% CI are shown in the brackets.

Moderation Analysis

The study also applied the meta-regression model for testing the moderation effects of student SES, grade level, centered year of publication, and publication type on the relationship between ATL and academic achievement. The moderation analysis was under the random-effects models because of the large number of heterogeneities. Table 2.4 shows the results. The results indicated that when combining multiple predictors into the meta-regression model, the significant positive effect of grade level only existed on the relation between ATL and reading among single timepoint designs. The publication year had a significant negative effect on the relation of ATL and mathematics achievement.

Table 2.4 Moderation analysis results

	Single timepoint designs		Longitudinal designs	
	Reading	Mathematics	Reading	Mathematics
SES	0.041	0.083	-0.076	-0.061
	(0.073)	(0.078)	(0.094)	(0.144)
Grade	0.095*	-0.013	0.014	0.059
	(0.047)	(0.051)	(0.085)	(0.085)
Publication type	-0.051	-0.101	-0.109	0.092
	(0.073)	(0.085)	(0.123)	(0.166)
Publication year	-0.008	-0.028*	-0.007	-0.014
	(0.007)	(0.009)	(0.017)	(0.043)

Note. Standard errors are shown in the parentheses. $p^{*} < .05$.

Discussion

The study applied four univariate meta-analysis to show the relationship between ATL and academic (reading and mathematics) achievement in childhood (preschool to elementary school) in the recent years from quasi-experimental evidence. The study reviewed 29 full-text studies in the final sample and included 21 studies into the meta-analysis. The studies investigating the ATL effect on reading performance were more than the studies testing the effect on mathematics performance. The studies exploring the concurrent relationships were more than the studies focusing on the long-term relationships. The weighed effect sizes under four conditions (two achievement × two quasi-experimental designs) were significantly different than zero. The range of the weighed effect sizes was from 0.328 to 0.366. The weighted mean effect sizes could be interpreted as medium effect sizes. The findings demonstrated that the relationship between ATL or learning-relative behaviors is positive and considerable in childhood. Also, the short-term and long-term effect both existed. The effect on reading achievement was stronger than the effect on mathematics achievement. The subgroup difference tests indicated that the weighed effect sizes were different in preschool, kindergarten and elementary school. However, when multiple predictors were taken into the same meta-regression model, the effect of grade level disappeared except for the relationship between ATL and concurrent reading achievement. The non-significant moderators showed that the effect of ATL on achievement was important for all students in childhood.

The present study also has some limitations. First, because ATL is a new domain compared with social-emotional learning and other traditional domains, a clear definition is still needed to define the components of ATL. A clear definition would help to collect studies and conduct future meta-analysis. Second, the univariate cases were applied in the current study to display clear results of the relationships, however, the univariate cases have to exclude several valuable studies which could not meet the selection criteria (e.g., the studies using combined achievement scores). Moreover, compared to interventions, quasi-experimental studies might have very different sample sizes and more complex modeling/ estimation approaches. Thus, the results from meta-analysis might not be robust. Future studies could work together with experts to make a clearer definition of ATL, extend participants age (e.g., middle school and college), and use proper research methods (e.g., multivariate meta-analysis) to get a more general conclusion of the relationship between ATL and academic achievement.

CHAPTER 3 POWER ANALYSIS IN META-ANALYSIS: A THREE-LEVEL MODEL

Introduction

The present chapter is methodologically oriented. It addressed an issue that could happen when conducting a meta-analysis. Specifically, this study focused on improving power analysis in meta-regression with hierarchical structures methodologically.

When conducting a meta-analysis, two weighted regression models are usually used in the statistical analysis. The two models are the fixed effects model and the random effects model. The fixed effects model assumes that there is one true population effect size, while the random effects model assumes that there is a variance from the systematic difference among studies. It captures a hierarchical structure that participants nested in the studies. Therefore, the random effects model is equivalent to the two-level model (Fernández-Castilla et al., 2020). However, in empirical research, a research group or a lab usually focuses on similar research topics. It is possible to collect several studies from the same research team in a meta-analysis. The protentional correlation of studies conducted by the same team or lab could influence the standard error of the weighted average effect size. Further, it could impact the calculation of the power statistics. Therefore, a meta-analysis with high power might be less credible due to a latent correlation between groups if between-group variance is ignored. The present study aims to introduce a three-level meta-regression model and explore the procedures to compute the power of weighted average effect size and moderators. Additionally, the study aims to show group-level variance potentially impacts the power statistics of the three-level meta-analysis regression model.

Literature Review

Quantitative research aims to draw statistical inferences about the population from limited samples. Researchers use inferential statistics and hypothesis testing to represent a population from sample data. Null hypothesis (H_0) and alternative hypothesis (H_A) are stated to display a research question, and then appropriate test statistics are applied to get the inference. A decision about whether to reject the null hypothesis depends on probability theory. The probability-related task examines the likelihood of observing the test statistics when assuming the null hypothesis is true. Researchers aim to reject the null hypothesis when the null hypothesis is false or retain the null hypothesis when the null hypothesis is true. However, because the decision is based on probability theory, a wrong decision is possibly made during the inference. Thus, keeping a small error in the inference decision is an important goal for conducting hypothesis testing.

There are two types of error in the hypothesis testing - Type I error and Type II error. Type I error, α, is the probability of rejecting the null hypothesis when it is true. In common, researchers set a significant level to limit Type I error. The critical significant level is usually 0.05. It indicates that the maximum probability of rejecting a true null hypothesis is 0.05. Type II error, β, is the probability of retaining a null hypothesis when it is false. In other words, it is the probability of not rejecting a null hypothesis when the alternative hypothesis is true. In empirical research, keeping a low Type I error and a low Type II error helps researchers make a correct and robust decision.

The power of a statistical test is referred to as the probability of finding a treatment effect when it exists (Cohen, 1977). The letter p is used to indicate power. Power represents the probability that a test correctly rejects the null hypothesis when it is false. Based on this definition, power can be calculated by $1 - \beta$, where β indicates the Type II error of the test. Power over 0.8 is usually considerable, indicating 80% chance of a real effect size stated in conclusion. Power could be influenced by significant level (refers to Type I error), sample size, variability in the measure of the response variable, and the effect size of the variable. Computing a prospective power is useful and important in experimental designs to determine how many subjects are needed to detect a treatment effect when it is true (Konstantopoulos, 2008). The studies with small power potentially surfer from low reproducibility of results and overestimated effects (Button et al., 2013).

A meta-analysis selects a pool of individual studies to detect an average effect size. Thus, meta-regression could increase statistical power to detect effects over what is obtained from individual studies because it involves more samples compared with one individual study (Miller & Pollock, 1994; Borenstein et al., 2021). Both prospective and retrospective statistical power for meta-analysis can be done with assumptions about the parameters in the specific meta-regression model (Valentine et al., 2010). Prospective statistical power could help researchers to determine how many studies need to be collected in a meta-analysis. Retrospective statistical power provides a measure to understand the risk level for a meta-analysis commits to type II error. Additionally, power analysis is more important in the meta-analysis than the analysis of a single study because such studies summarize similar research and influence the theory and practice of the field strongly (Cafri et al., 2010). However, the number of studies could not always increase the statistical power. Several components influence power. Results from a meta-analysis should be interpreted with great care. Therefore, finding an unbiased power of a meta-analysis is critical to measure a good meta-analysis study.

The existing methodology regarding power analysis for meta-regression can be used for both fixed and random effects models (Hedges & Pigott, 2001). The researchers also developed power statistics for the heterogeneity (or variation) test of effect size parameters across studies. Also, previous studies have considered the power analysis for moderators in meta-regression models (Hedges & Pigott, 2004). Thus far, power analysis for random effects models in meta-regression has focused on two-level models where studies are at the second level. However, more complicated data structures exist in empirical meta-analysis. A natural extension of that work is to extend the methods for random effects models where a third level (e.g., research teams/labs) is added into the model.

In an empirical systematic review and meta-analysis, the final sample studies are possible from the same research groups or research labs. In this case, the studies included in the meta-analysis have a dependency because they are nested within research groups or labs. If there is a dependent effect size problem in a meta-analysis, using two-level meta-regression likely underestimates the standard error. This additional dependency needs to be taken into account in calculating power because ignoring heterogeneity between groups possibly influences statistical power.

There are three ways to account for the dependent effect size issue - ignoring dependence, avoiding dependence, and modeling dependence. Under the ignoring dependence strategy, researchers ignore the potential dependence among studies in the meta-analysis. However, this strategy is inappropriate because the existing dependency might lead to bias in the following estimations. Under the avoiding dependence strategy, one way is to choose one effect size for each study. Another way is choosing effect size based on the units of analysis, for instance, choosing one effect size from each sample, each research group, or each study.

However, it is hard for researchers to decide which one should be included in their meta-regression model. Another common strategy of deciding effect size within a unit is to average effect size in each unit (Van den Noortgate et al., 2013). However, using average effect size will reduce the variance among studies. Therefore, compared with other strategies, modeling dependence is a better way to deal with potential heterogeneity between groups.

One way to resolve this issue under modeling dependence strategy is to use three-level meta-regression models (Konstantopoulos, 2011; Van den Noortgate et al., 2013). A three-level meta-analytic model (including power analysis) assumes that the between-group variance is not zero, which indicates that studies are nested in research groups or labs. The three-level meta-regression model shows several advantages to model between-group variance. First, it is a very flexible model because it could account for several sources of dependence at the same time. Second, it is a relatively intuitive and straightforward way to account for dependence. Additionally, it automatically accounts for the hierarchical structure in the data (Van den Noortgate et al., 2013).

Present Study

False accounting potential group dependence leads to biased power statistics in the meta-regression model. To address this issue, the present study extended the work on power analysis for the two-level random effects model to the three-level model where studies were at the second level and research teams/labs are at the third level (Konstantopoulos, 2011). A three-level model would provide more accurate estimates of power under the assumption that variability between research teams is not negligible. The present study aimed to fill in that gap in the literature to figure out the power of the three-level meta-regression model. In details, the research questions are:

(1) How to calculate the power of the statistical test for weighted average effect size in a three-level meta-analytic model?

(2) How to calculate the power of the statistical test for moderators in a three-level meta-regression model?

(3) How could the third level (group-level) heterogeneity affect statistical power of weighted average effect size in meta-analysis from a simulation study?

The significances of the study are listed here. First, the study was a methodological development of power analysis in the meta-analysis by developing the formulas for power statistics in three-level model. Second, it considered more complicated data structures in meta-analysis and provides unbiased powers measure in the three-level model. Third, the study provided evidence about how group-level heterogeneity affects statistical power in meta-analysis.

Statistical Modeling

Power in Two-Level Meta-Regression Models (Intercept Only)

The two-level meta-regression model is equivalent to the random effects meta-regression model, which assumes effect sizes are nested in studies. It considers the amount of heterogeneity observed among effect sizes across studies (Hedges & Vevea, 1998; Hedges & Olkin, 2014).

Power calculation in a two-level meta-regression model has been shown in Hedges and Pigott's work (2001). How statistical power relates to a weighted average effect size ($\bar{T}_{\cdot}^{*}$), the effect size assumed in the null hypothesis (T_0), Type I error (α), and the standard error ($SE(\bar{T}_{\cdot}^{*})$) of the weighted average effect size in a random effects meta-regression model is generally shown as

$$p \propto \frac{(\bar{T}_{\cdot}^{*} - T_0) \cdot \alpha}{SE(\bar{T}_{\cdot}^{*})}. \tag{3.1}$$

It indicates the statistical power could be increased by a larger weighted pooled effect size, a higher significance level (Type I error), or a smaller standard error of the weighted pooled effect size.

The assumptions of a two-level model are 1) there is heterogeneity of the sampling error because the sample sizes of studies are usually different; 2) random effects are distributed identically at the between-study level; 3) Individuals are independent of each other, which indicates no correlation between error terms at the first level; 4) Studies are independent of each other, which means no correlation between error terms at the second level. Therefore, in a two-level model, the variance-covariance matrix of error term could be written as

$$\boldsymbol{V}^{*} = Diag(\tau^2 + v_1, \tau^2 + v_2, \ldots, \tau^2 + v_k) = \begin{bmatrix} \tau^2 + v_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \tau^2 + v_k \end{bmatrix}, \tag{3.2}$$

which is introduced briefly in Chapter 2.

Power is calculated under the distribution when the alternative hypothesis of the study is true. Thus, it follows a non-central distribution. A non-centrality parameter (λ) needs to be detected for calculating the following probabilities (power statistics). The non-centrality parameter can be obtained by substituting the sample estimates with the population parameters in the formula with a Z test. In a random effects meta-regression model, the null hypothesis is the weighted average effect size is zero. The non-centrality parameter λ^{*} can be calculated using

$$\lambda^* = \frac{\bar{T}_\cdot^* - 0}{\sqrt{v_\cdot^*}} = \frac{\frac{\sum_{i=1}^{k} w_i T_i}{\sum_{i=1}^{k} w_i} - 0}{\sqrt{\frac{1}{\sum_{i=1}^{k} w_i}}} = \frac{\sum_{i=1}^{k} w_i T_i}{\sqrt{\sum_{i=1}^{k} w_i}}, \tag{3.3}$$

where $w_i = \frac{1}{v_i + \tau^2}$.

Thus, the non-centrality parameter is computed by the weighted average effect size ($\bar{T}_\cdot^*$) and sampling variance of the random effects estimate ($v_\cdot^*$). Equation (3.3) shows how to compute the non-centrality parameter. The numerator is the sum of the product of weight and effect size in each study and the denominator is the square root of the sum of weights from each study. To note that the non-centrality parameter is resulted as a scalar.

Typically, power of a two-tailed test is usually computed in empirical studies. Thus, after getting the non-centrality parameter, the statistical power in a two-tailed Z test can be expressed as Equation (3.4)

$$p = 1 - \beta = P[\,|Z'(\lambda^*)| \geq Z_0] = 1 - \Phi(1.96 - \lambda^*) + \Phi(-1.96 - \lambda^*), \tag{3.4}$$

which is to calculate the probability of rejecting the null hypothesis when the null hypothesis is false. $Z'(\lambda^*)$ indicates the λ^*'s percent point of the distribution. $\Phi(x)$ indicates a standard normal distribution cumulative distribution function (cdf). When setting type I error α equals to 0.05, the critical value of the distribution, Z_0, is 1.96 for a two-tailed Z test.

Additionally, the statistical power in a one-tailed Z test can be expressed as

$$p = 1 - \beta = P[\,Z'(\lambda^*) \geq Z_0] = 1 - \Phi(1.65 - \lambda^*), \tag{3.5}$$

when setting type I error α equals to 0.05, the critical value of the distribution, Z_0, is 1.65 for a one-tailed Z test.

To illustrate a case of computing power of mean effect size from a two-level meta-regression model in practice, the present study shows a sample example here. We suppose a meta-analysis has ten studies with different effect sizes and within-study variances. Between-study variances are same for all studies under the two-level model assumption. Therefore, we could compute a specific weight for each study in the sample. The parameters are shown in Table 3.1.

Table 3.1 An illustrated two-level meta-analysis sample with intercept only

Study ID	Effect size (T_i)	Within-study variance (v_i)	Between-study variance (τ^2)	Weight (w_i)
1	0.42	0.13	0.05	5.56
2	0.27	0.12	0.05	5.88
3	0.28	0.08	0.05	7.69
4	0.41	0.10	0.05	6.67
5	0.46	0.11	0.05	6.25
6	0.32	0.13	0.05	5.56
7	0.30	0.16	0.05	4.76
8	0.34	0.07	0.05	8.33
9	0.54	0.12	0.05	5.88
10	0.39	0.19	0.05	4.17

We follow Equation (3.3) to get the non-centrality parameter

$$\lambda^* = \frac{\sum_{i=1}^{10} w_i T_i}{\sqrt{\sum_{i=1}^{10} w_i}} = 2.89,$$

where 10 studies are in the example. And the non-centrality parameter is 2.89.

Then, we put this number into Equation (3.4) to compute the power in the case. The formula is shown as

$$p = 1 - \Phi(1.96 - 2.89) + \Phi(-1.96 - 2.89) = 0.82.$$

The power of weighted average effect size in the example is 0.82 in a two-tailed Z test, which consider as a good power the weighted average effect size in meta-analysis.

Power for Moderators in Two-Level Meta-Regression Models

Hedges and Pigott (2004) developed a method to calculate statistical power in moderation analysis in two-level (random-effects) meta-regression models. Moderators are at study level because they represent the differences among studies. The observed effect with *p* moderators in a within-study model and in a between-study model could be written as

$$T_i = \theta_i + \varepsilon_i,\ where\ \varepsilon_i \sim N(0, v_i), \tag{3.6}$$

$$\theta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \zeta_i,\ where\ \zeta_i \sim N(0, \tau^2). \tag{3.7}$$

Combining with the components from both levels, a general equation in a single level for *p* moderators in a two-level meta-regression model is shown as

$$T_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \eta_i\ ,\ where\ \eta_i \sim N(0, \tau^2 + v_i), \tag{3.8}$$

Where the error term η_i follows a normal distribution with mean equals to 0 and variance equals a sum of within-study variance (v_i) and between-study variance (τ^2).

Equation (3.8) could be written to a matrix notation as $\boldsymbol{T} = \boldsymbol{X\beta} + \boldsymbol{\eta}$, where $\boldsymbol{\eta}$ has a *k* variate normal distribution with mean 0 and variance-covariance matrix $\boldsymbol{V}$, if *k* studies are included in the meta-regression model. To note that variables and parameters are in boldface indicate vectors in the equations. By using generalized least square (GLS) estimation method, the estimated coefficients of the moderators ($\widehat{\boldsymbol{\beta}}^*$) and the variance of the estimated moderators ($Var(\widehat{\boldsymbol{\beta}}^*)$) are solved in

$$\widehat{\boldsymbol{\beta}}^* = [\boldsymbol{X}'(\boldsymbol{V}^*)^{-\mathbf{1}}\boldsymbol{X}]^{-\mathbf{1}}\ \boldsymbol{X}'(\boldsymbol{V}^*)^{-\mathbf{1}}\boldsymbol{T} = (\textstyle\sum_{\boldsymbol{i=1}}^{\boldsymbol{k}} \boldsymbol{X}_{\boldsymbol{i}}'\ \boldsymbol{W}\boldsymbol{X}_{\boldsymbol{i}})^{-1} \textstyle\sum_{\boldsymbol{i=1}}^{\boldsymbol{k}} \boldsymbol{X}_{\boldsymbol{i}}'\boldsymbol{W}\boldsymbol{T}_{\boldsymbol{i}}, \tag{3.9}$$

$$Var(\widehat{\boldsymbol{\beta}}^*) = \boldsymbol{\Sigma}^* = [\boldsymbol{X}'(\boldsymbol{V}^*)^{-\mathbf{1}}\boldsymbol{X}]^{-\mathbf{1}} = (\textstyle\sum_{\boldsymbol{i=1}}^{\boldsymbol{k}} \boldsymbol{X}_{\boldsymbol{i}}'\ \boldsymbol{W}\boldsymbol{X}_{\boldsymbol{i}})^{-1}, \tag{3.10}$$

where $\boldsymbol{V}^*$ is the variance-covariance matrix from Equation (3.2) and $\boldsymbol{W} = (\boldsymbol{V}^*)^{-\mathbf{1}}$ is the weight matrix.

Different methods could be used to estimate the between-study variance component (τ^2). Hedges and Pigott (2004) used the same way for computing variance components in ANOVA. Other popular ways include methods such as the method of moments (MOM), maximum likelihood estimation (MLE), and restricted maximum likelihood estimation (RMLE) (Langan et al., 2019). MLE aims to solve the parameters to maximize the likelihood function of the data (Corbeil & Searle, 1976). It could provide simultaneous estimations of the fixed effects and the variance components in multilevel regression. It assumes fixed effects are known when estimating the variance components. Iterations might be required to get the estimations, such as an expectation-maximization (EM) algorithm or a fisher scoring algorithm (Raudenbush & Bryk, 2002). RMLE is less biased than MLE when the cluster size is small. Differing from the estimation procedures in MLE, RMLE estimates the fixed effects when estimating the variances (Peugh, 2010; Boedeker, 2017). Veroniki and co-researchers (2015) identified over ten estimators of the between-study variance in meta-analysis models and suggested that RMLE was the better estimator for continuous outcomes. It tends to outperform the alternatives in the simulation studies (Langan et al., 2019). RMLE leads to the use in multilevel regression software packages, such as HLM8 (Raudenbush et al., 2019).

The null hypothesis in the case is there is no relationship between moderator *j* and effect size ($\beta_j = 0$). Thus, the non-centrality parameter in the test can be computed by

$$Z_j^* = \frac{\hat{\beta}_j^* - 0}{\sqrt{var(\hat{\beta}_j^*)}}, \tag{3.11}$$

where $var(\hat{\beta}_j^*)$ is the variance of $\hat{\beta}_j^*$ given by the *j*th diagonal element of the matrix $\boldsymbol{\Sigma}^*$.

Then the power for test of individual regression coefficients (the coefficients of moderators) could be calculated by

$$p = 1 - \beta = P[\,|Z'(\lambda^*)| \geq Z_0] = 1 - \Phi\left(1.96 - Z_j^*\right) + \Phi\left(-1.96 - Z_j^*\right), \qquad (3.12)$$

$$p = 1 - \beta = P[\,Z'(\lambda^*) \geq Z_0] = 1 - \Phi\left(1.65 - Z_j^*\right). \qquad (3.13)$$

The power shown in Equation (3.12) is for a two-tailed Z test and Equation (3.13) is for a one-tailed Z test where the type I error is set to 0.05.

To illustrate a case of computing power of the moderators from a two-level meta-regression model in practice, the present study shows a sample example here. We use the meta-analysis sample from previous section and suppose the study has one moderator (X_1) at study level. For example, the moderator is a categorical variable with three categories and the categories are randomly assigned to the stud in the example. The parameters are shown in Table 3.2.

Table 3.2 An illustrated two-level meta-analysis sample with one moderator

Study ID	Effect size (T_i)	Within-study variance (v_i)	Between-study variance (τ^2)	Weight (w_i)	Moderator (x_{1i})
1	0.42	0.13	0.05	5.56	1
2	0.27	0.12	0.05	5.88	3
3	0.28	0.08	0.05	7.69	1
4	0.41	0.10	0.05	6.67	2
5	0.46	0.11	0.05	6.25	2
6	0.32	0.13	0.05	5.56	3
7	0.30	0.16	0.05	4.76	1
8	0.34	0.07	0.05	8.33	1
9	0.54	0.12	0.05	5.88	3
10	0.39	0.19	0.05	4.17	2

Therefore, we could compute the regression coefficient of the moderator and its variance using Equation (3.9) and (3.10). The equations are

$$\hat{\beta}_1^* = (\sum_{i=1}^{10} w_i\, x_{1i}^2)^{-1} \sum_{i=1}^{10} w_i\, x_{1i} t_i = 0.17,$$

$$Var(\hat{\beta}_1^*) = (\sum_{i=1}^{10} w_i\, x_{1i}^2)^{-1} = 0.004.$$

The coefficient of the moderator is 0.17 and the variance is 0.004 in the example.

Then, we follow Equation (3.11) to get the non-centrality parameter:

$$Z_1^* = \frac{0.17}{\sqrt{0.004}} = 2.71.$$

The non-centrality parameter is 2.71 in the illustrated example. Then, we put this number into Equation (3.12) to compute the power in the case. The computation is shown as

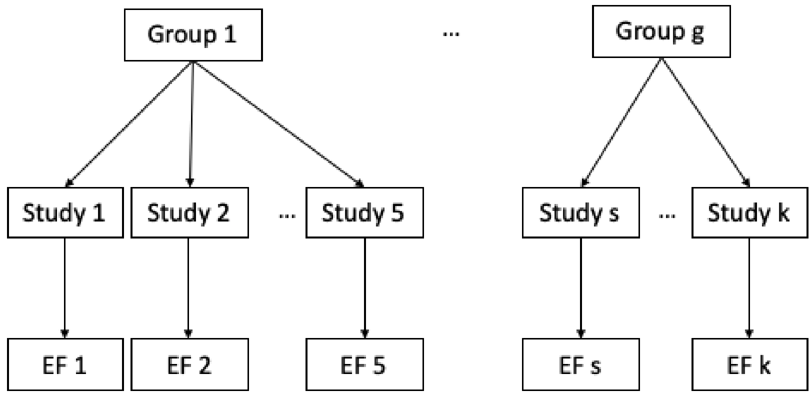
$$p = 1 - \Phi(1.96 - 2.71) + \Phi(-1.96 - 2.71) = 0.77.$$

The power of the moderator in the example is 0.77 in a two-tailed Z test, which consider as a fair power for moderation analysis in a meta-analysis.

Power in Three-Level Meta-Regression Models (Intercept Only)

To compute statistical power in the three-level meta-regression model with intercept only, building a three-level model is necessary. The study first focused on an unconditional model, which means no predictors at any level. The power of the weighted average effects size is tested in the case. The intercept of the study level (level-2) is random at the group level (level-3). Working with a simple case would be helpful to illustrate the main ideas of the present studies. The study uses a univariate case, which means each study in the model only has one effect size. Figure 3.1 illustrates the hierarchical structure of the three-level model with *g* groups, *k* studies and *k* effect sizes.

Figure 3.1 An illustrated structure of a univariate case with three levels



The model with three levels could be written as

$$\text{Level-1 effect size level: } T_{ig} = \vartheta_{ig} + \varepsilon_{ig},\ \varepsilon_{ig} \sim N(0, v_i), \tag{3.14}$$

$$\text{Level-2 study level: } \vartheta_{ig} = \beta_{0g} + \eta_{ig}, \eta_{ig} \sim N(0, \tau_{(2)}^2), \tag{3.15}$$

$$\text{Level-3 group level: } \beta_{0g} = \gamma_{00} + \nu_{0g}, \nu_{0g} \sim N(0, \tau_{(3)}^2). \tag{3.16}$$

At the first level, effect size level, an observed effect size (T_{ig}) of study *i* in group g is a sum of an effect size (ϑ_{ig}) and a within-study error (ε_{ig}). The within-study error follows a normal distribution with mean 0 and variance (v_i). At the second level, study level, the effect size (ϑ_{ig}) from participant level equals to an effect size (β_{0g}) plus a between-study error (η_{ig}). The between-study error follows a normal distribution with mean 0 and variance $\tau_{(2)}^2$. At the third level, group level, the effect size (β_{0g}) from study level equals to a true effect size (γ_{00}) plus a between-group error (ν_{0g}). The between-group error follows a normal distribution with mean 0 and variance $\tau_{(3)}^2$.

All three levels are written in a single level notation as

$$T_{ig} = \gamma_{00} + \nu_{0g} + \eta_{ig} + \varepsilon_{ig}. \tag{3.17}$$

It shows the observed effect size of a study is a sum of a true effect size and three parts of error - within-study error, between-study error and between-group error.

The next step is to construct the structure of the variance-covariance matrix of error term for the three-level meta-regression model. It is important for detecting the structure of the variance-covariance matrix because the inverse of the matrix would be used as a weight matrix in the following steps for computing the weighted average effect size and its variance. Also, the wight matrix further influences the power statistics. When introducing the third level (group level) into the model, the variance-covariance matrix of error $V_{(3,g)}$ in group g becomes the sum of the diagonal matrix in the two-level model and a matrix with element $\tau_{(3)}^2$ everywhere. The matrix structure is shown as

$$V_{(3,g)} = \begin{bmatrix} \tau_{(2)}^2 + v_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \tau_{(2)}^2 + v_k \end{bmatrix} + \begin{bmatrix} \tau_{(3)}^2 & \cdots & \tau_{(3)}^2 \\ \vdots & \ddots & \vdots \\ \tau_{(3)}^2 & \cdots & \tau_{(3)}^2 \end{bmatrix}$$

$$= \begin{bmatrix} \tau_{(3)}^2 + \tau_{(2)}^2 + v_1 & \cdots & \tau_{(3)}^2 \\ \vdots & \ddots & \vdots \\ \tau_{(3)}^2 & \cdots & \tau_{(3)}^2 + \tau_{(2)}^2 + v_k \end{bmatrix}. \quad (3.18)$$

The underlying assumption is groups are independent of each other but studies in the same group have correlations. Thus, the variance of study *i* in group g is $\tau_{(3)}^2 + \tau_{(2)}^2 + v_i$, and the covariance for studies in same group is $\tau_{(3)}^2$. The variance $\tau_{(3)}^2$ captures the dependency of outcomes with groups. The variance-covariance matrix of group g follows Konstantopoulos's notation (2011) could be written as

$$V_{(3,g)} = V_{(2,g)} + 1_{(3,g)} T_3 1_{(3,g)}^T = I_{n_g} \otimes \{\tau_{(2)}^2 + v_i\} + \tau_{(3)}^2 1_{n_g} 1_{n_g}^T, \quad (3.19)$$

where $V_{(2,g)}$ is the variance-covariance matrix of a two-level model, $1_{(3,g)}$ is a vector of ones, T_3 is a matrix of random effects at the group level, n_g indicates the number of studies in group *g*, I_{n_g} is an $n_g \times n_g$ identity matrix, and $1_{(n_g)}$ is a vector of n_g ones.

The methods to estimate the variance components are same to the two-level model, such as MLE and RMLE. For instance, the full log-likelihood function for group *g* of the three-level model is:

$$L_g\left(v_g, \tau_{(2)}^2, \tau_{(3)}^2\right) = -\frac{n_g}{2}\log(2\pi) - \frac{1}{2}\log\left|V_{(3,g)}\right| - \frac{1}{2}e_g' V_{(3,g)}^{-1} e_g. \tag{3.20}$$

where $e_g = T_g - \gamma_{00}$ is the sum of error terms in group *g*, $\left|V_{(3,g)}\right|$ indicates the determinant of $V_{(3,g)}$. The sampling variance (v_i) within studies is usually assumed fixed and known in meta-analysis. Because groups are independent of each other, the log-likelihood for entire model is the sum of unit log-likelihoods in Equation (3.20). The estimated variances could be gained when maximizing the log-likelihood function of the entire model.

Overall, the whole variance-covariance matrix for a three-level meta regression with *k* studies nested in *m* groups is a block matrix with *m* matrices on the diagonal line. Suppose in the first group we have *t* studies and the last group we have *s* studies, the illustrated variance-covariance matrix $\boldsymbol{V_3}$ is shown as

$$\boldsymbol{V_3} = \begin{bmatrix} \begin{bmatrix} \tau_{(3)}^2 + \tau_{(2)}^2 + v_1 & \cdots & \tau_{(3)}^2 \\ \vdots & \ddots & \vdots \\ \tau_{(3)}^2 & \cdots & \tau_{(3)}^2 + \tau_{(2)}^2 + v_t \end{bmatrix} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \begin{bmatrix} \tau_{(3)}^2 + \tau_{(2)}^2 + v_{k-s+1} & \cdots & \tau_{(3)}^2 \\ \vdots & \ddots & \vdots \\ \tau_{(3)}^2 & \cdots & \tau_{(3)}^2 + \tau_{(2)}^2 + v_k \end{bmatrix} \end{bmatrix}. \tag{3.21}$$

Also, it could be written as

$$\boldsymbol{V_3} = Diag\left(V_{(3,1)}, V_{(3,2)}, \dots, V_{(3,m)}\right) = I_m \otimes \left\{V_{(3,g)}\right\}$$

$$= I_m \otimes \left\{V_{(2,g)} + 1_{(3,g)} T_3 1_{(3,g)}^T\right\}, \tag{3.22}$$

where {} indicates the matrices in each group.

The inverse of the variance-covariance matrix is used as a weight matrix into the generalized least square estimation. The inverse of the block matrix equals to the inverse of each block in the matrix, which could be written as

$$\boldsymbol{V_3^{-1}} = I_m \otimes \left\{V_{(3,g)}^{-1}\right\} = I_m \otimes \begin{bmatrix} (V_{(2,1)} + 1_{(3,1)} T_3 1_{(3,1)}^T)_1^{-1} \\ \vdots \\ (V_{(2,m)} + 1_{(3,m)} T_3 1_{(3,m)}^T)_m^{-1} \end{bmatrix}. \tag{3.23}$$

To note that a block-diagonal matrix is invertible if and only if the blocks on the diagonal are invertible. By using the standard results (Longford, 1987; Konstantopoulos, 2011), the inverse of $V_{(3,g)}$ could be separated as:

$$V_{(3,g)}^{-1} = V_{(2,g)}^{-1} - V_{(2,g)}^{-1} 1_{(3,g)} (T_3^{-1} + 1_{(3,g)}^T V_{(2,g)}^{-1} 1_{(3,g)})^{-1} 1_{(3,g)}^T V_{(2,g)}^{-1}. \tag{3.24}$$

The non-centrality parameter λ_3^* in the three-level meta-regression model with no predictors could be calculated by using the weighted average effect size and the variance of the weighted average effect size, which is shown in

$$\lambda_3^* = \frac{\left[\boldsymbol{1}'(\boldsymbol{V_3})^{-\boldsymbol{1}}\boldsymbol{1}\right]^{-\boldsymbol{1}} \boldsymbol{1}'\boldsymbol{V_3^{-1}}\boldsymbol{T} - \boldsymbol{0}}{\sqrt{\left[\boldsymbol{1}'(\boldsymbol{V_3})^{-\boldsymbol{1}}\boldsymbol{1}\right]^{-\boldsymbol{1}}}} = \frac{\boldsymbol{1}'\boldsymbol{W_3}\boldsymbol{T}}{\sqrt{\boldsymbol{1}'\boldsymbol{W_3}\boldsymbol{1}}}, \tag{3.25}$$

where $\boldsymbol{1}$ is a vector of ones and $\boldsymbol{W_3} = (\boldsymbol{V_3})^{-\boldsymbol{1}}$ is the weight matrix in the case, $\boldsymbol{T}$ is the vector of observed effect sizes. To note that, the numerator is the sum of products of the weight and the effect size in each study from each group and the denominator is the square root of the sum of the weight in each study from each group (all elements in the weight matrix), which could be written as

$$\lambda_3^* = \frac{\Sigma_{g=1}^{m} \Sigma_{t=1}^{k_g} (\Sigma_{s=1}^{k_g} W_{3(st,g)}) T_{(t,g)}}{\sqrt{\Sigma_{g=1}^{m} \Sigma_{t=1}^{k_g} \Sigma_{s=1}^{k_g} W_{3(st,g)}}}, \tag{3.26}$$

where k_g is the number of studies in the *g*th group, *m* is the number of groups, $W_{3(st,g)}$ indicates the element at the *s*th row and *t*th column in the *g*th group from the weight matrix, and $T_{(t,g)}$ indicates the *t*th effect size. And finally, the non-centrality parameter is resulted as a scalar because the numerator and the denominator are both scalar. Therefore, to get the power statistics in the three-level model with no predictors at the second and the third level, we can put λ_3^* into Equation (3.27) and (3.28) for a two-tailed Z test and a one-tailed Z test when the type I error is set to 0.05:

$$p = 1 - \beta = P[\,|Z'(\lambda_3^*)| \geq Z_0] = 1 - \Phi(1.96 - \lambda_3^*) + \Phi(-1.96 - \lambda_3^*), \tag{3.27}$$

$$p = 1 - \beta = P[\,Z'(\lambda_3^*) \geq Z_0] = 1 - \Phi(1.65 - \lambda_3^*). \tag{3.28}$$

To illustrate a case of computing power of the mean effect size from a three-level meta-regression model in practice, the present study shows a sample example here. We continue to use the meta-analysis sample from previous sections. The difference in the case is that between-group variance is introduced into the analysis. Therefore, we need to compute a specific weight for each study to capture within-study variance, between-study variance and between-group variance. The weights are different from the weight from Table 3.1. In this case, we assume ten studies come from three research groups and between-group variance equals to 0.02. The parameters are shown in Table 3.3.

Table 3.3 An illustrated three-level meta-analysis sample with intercept only

Study ID	Group ID	Effect size (T_{ig})	Within-study variance (v_i)	Between-study variance ($\tau^2_{(2)}$)	Between-group variance ($\tau^2_{(3)}$)
1	1	0.42	0.13	0.05	0.02
2	2	0.27	0.12	0.05	0.02
3	2	0.28	0.08	0.05	0.02
4	3	0.41	0.10	0.05	0.02
5	1	0.46	0.11	0.05	0.02
6	3	0.32	0.13	0.05	0.02
7	3	0.30	0.16	0.05	0.02
8	3	0.34	0.07	0.05	0.02
9	1	0.54	0.12	0.05	0.02
10	2	0.39	0.19	0.05	0.02

The weighted matrix could be constructed using within-study variance, between-study variance and between-group variance as

$$\boldsymbol{W_3} = (\boldsymbol{V_3})^{-1} = \begin{bmatrix} \begin{bmatrix} 0.20 & 0.02 & 0.02 \\ 0.02 & 0.18 & 0.02 \\ 0.02 & 0.02 & 0.19 \end{bmatrix} & \cdots & 0 \\ \vdots & \begin{bmatrix} 0.19 & 0.02 & 0.02 \\ 0.02 & 0.15 & 0.02 \\ 0.02 & 0.02 & 0.26 \end{bmatrix} & \vdots \\ 0 & \cdots & \begin{bmatrix} 0.17 & 0.02 & 0.02 & 0.02 \\ 0.02 & 0.20 & 0.02 & 0.02 \\ 0.02 & 0.02 & 0.23 & 0.02 \\ 0.02 & 0.02 & 0.02 & 0.14 \end{bmatrix} \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} \begin{bmatrix} 5.10 & -0.51 & -0.48 \\ -0.51 & 5.67 & -0.54 \\ -0.48 & -0.54 & 5.27 \end{bmatrix} & \cdots & 0 \\ \vdots & \begin{bmatrix} 5.37 & -0.67 & -0.36 \\ -0.67 & 6.82 & -0.47 \\ -0.36 & -0.47 & 3.91 \end{bmatrix} & \vdots \\ 0 & \cdots & \begin{bmatrix} 6.08 & -0.49 & -0.42 & -0.74 \\ -0.49 & 5.15 & -0.35 & -0.61 \\ -0.42 & -0.35 & 4.46 & -0.53 \\ -0.74 & -0.61 & -0.52 & 7.41 \end{bmatrix} \end{bmatrix}.$$

Each block in the matrix indicates one group from the example. All numbers are round to two decimals.

Thus, we could use the weight matrix in Equation (3.26) to compute the non-centrality parameter. The computation results are shown as

$$\lambda_3^* = \frac{\Sigma_{g=1}^{m}\Sigma_{t=1}^{k_g}(\Sigma_{s=1}^{k_g} W_{3(st,g)})T_{(t,g)}}{\sqrt{\Sigma_{g=1}^{m}\Sigma_{t=1}^{k_g}\Sigma_{s=1}^{k_g} W_{3(st,g)}}} = \frac{15.979}{\sqrt{42.968}} = 2.44.$$

The non-centrality parameter is 2.44 in this example. Compared with the non-centrality parameter of the weighted average effect size in two-level meta-regression model, the current non-centrality parameter is smaller. The reason is the variation from the third level is considered into the model.

Then, we put this number into Equation (3.27) to compute the power in the case. The equation is

$$p = 1 - \Phi(1.96 - 2.44) + \Phi(-1.96 - 2.44) = 0.69.$$

The power of weighted average effect size in the example is 0.69 in a two-tailed Z test. The power decreases when the third level (group level) is introduced into the model. The result from the illustrated example demonstrates that the group-level variance could impact the power of weighted average effect size in a meta-analysis.

Power for Moderators in Three-Level Meta-Regression Models

When the study aims to test the moderation effects in a meta-analysis, the calculations of power for moderators in a three-level meta-regression model have similar steps to the previous section. Frist, we need to find the estimated coefficients (moderators in level-2 and level-3). Second, the variances of those coefficients need to be detected. In details, the structure of the variance-covariance matrix in a three-level meta-analysis with moderators at level-2 and level-3 should be figured out. For each moderator, the null hypothesis is there is no moderation effect on effect size. Then we could calculate the non-centrality parameter in the alternative distribution and uses it to detect statistical power for the moderators in a three-level meta-regression model.

Moderators with No Random Effects

The present study follows the procedures of calculating statistical power for moderators in two-level mate-regression model to extend the calculation of power for moderators in three-level meta-regression model. A three-level meta-regression with *p* moderators in level-2 and *q* moderators in level-3 is shown in

$$\text{Level-1 effect size level: } T_{ig} = \vartheta_{ig} + \varepsilon_{ig},\ \varepsilon_{ig} \sim N(0, v_i), \tag{3.29}$$

$$\text{Level-2 study level: } \vartheta_{ig} = \beta_{0g} + \beta_{1g} x_{1ig} + \cdots + \beta_{pg} x_{pig} + \eta_{ig},\ \eta_{ig} \sim N(0, \tau_{(2)}^2), \tag{3.30}$$

$$\text{Level-3 group level: } \beta_{0g} = \gamma_{00} + \gamma_{01} z_{1g} + \cdots + \gamma_{0q} z_{qg} + \nu_{0g}, \nu_{0g} \sim N\left(0, \tau_{(3)}^2\right), \tag{3.31}$$

where *x* and *z* are moderators at level-2 and level-3. In the current model, only the intercept in the second level is random at the third level. All other level-2 slopes are fixed at level-3, namely $\beta_{lg} = \gamma_{lg}$, where *l* indicates the *l*th slope and *g* indicates the *g*th group.

The above equations could be written in a single-level equation as:

$$T_{ig} = \gamma_{00} + \boldsymbol{X}_p \mathbf{B}_p + \boldsymbol{Z}_q \boldsymbol{\Gamma}_q + \nu_{0g} + \eta_{ig} + \varepsilon_{ig} \tag{3.32}$$

where ***X*** and ***Z*** indicate two vectors of moderators When the slopes (except the intercept) at level-2 are fixed at level-3, the variance-covariance matrix of error is same to the matrix in Equation (3.21) because no extra random effects need to be estimated in the model. Thus, we could use the inversed matrix as the weight matrix to estimate slopes (regression coefficients) in Equation (3.33) and their variances in Equation (3.34). The formulas are shown as

$$\begin{bmatrix} \widehat{\mathbf{B}}^{\dagger} \\ \widehat{\boldsymbol{\Gamma}}^{\dagger} \end{bmatrix} = [(\boldsymbol{X} + \boldsymbol{Z})'(\boldsymbol{V}_{\mathbf{3}})^{-\mathbf{1}}(\boldsymbol{X} + \boldsymbol{Z})]^{-\mathbf{1}}\ (\boldsymbol{X} + \boldsymbol{Z})'(\boldsymbol{V}_{\mathbf{3}})^{-\mathbf{1}}\boldsymbol{T}, \tag{3.33}$$

$$Var(\widehat{\mathbf{B}}^{\dagger},\widehat{\mathbf{\Gamma}}^{\dagger}) = \boldsymbol{\Sigma}^{\dagger} = [(\boldsymbol{X}+\boldsymbol{Z})'(\boldsymbol{V_3})^{-\mathbf{1}}(\boldsymbol{X}+\boldsymbol{Z})]^{-\mathbf{1}}. \quad (3.34)$$

To note that, $(\boldsymbol{V_3})^{-\mathbf{1}}$ could be written as $\boldsymbol{W_3}$, which is the weight matrix.

Further, the non-centrality parameters could be computed by using the estimated coefficients and their correspondent variances. For instance, the non-centrality parameters $\lambda_{3,l}^{\dagger}$ for moderator *l* at the third level could be calculated in the model in

$$\lambda_{3,l}^{\dagger} = \frac{\hat{\gamma}_l^{\dagger} - 0}{\sqrt{Var(\hat{\gamma}_l^{\dagger})}}. \quad (3.35)$$

We can put $\lambda_{3,l}^{\dagger}$ into Equation (3.36) and (3.37) for a two-tailed Z test and a one-tailed Z test when the type I error is set to 0.05. Thus, the power of the present moderator *l* in the three-level meta-regression model could be obtained in:

$$p = 1 - \beta = P\left[\,\left|Z'\left(\lambda_{3,l}^{\dagger}\right)\right| \geq Z_0\right] = 1 - \Phi\left(1.96 - \lambda_{3,l}^{\dagger}\right) + \Phi\left(-1.96 - \lambda_{3,l}^{\dagger}\right), \quad (3.36)$$

$$p = 1 - \beta = P\left[\,Z'\left(\lambda_{3,l}^{\dagger}\right) \geq Z_0\right] = 1 - \Phi\left(1.65 - \lambda_{3,l}^{\dagger}\right). \quad (3.37)$$

To illustrate a case of computing power of the moderators from a three-level meta-regression model in practice, the present study shows a sample example here. We continue to use the meta-analysis sample from previous sections. The example uses the parameters from Table 3.3. And the level-2 moderator is still the same moderator from Table 3.2. The present case also assumes a categorical moderator at the third level (Table 3.4). The weight matrix is the same weight matrix in the last example, because there is no random effect of the level-2 moderator.

Table 3.4 An illustrated three-level meta-analysis sample with moderators

Study ID	Group ID	Effect size (T_{ig})	Within-study variance (v_i)	Between-study variance ($\tau^2_{(2)}$)	Between-group variance ($\tau^2_{(3)}$)	Moderator level-2 (x_{1i})	Moderator level-3 (z_1)
1	1	0.42	0.13	0.05	0.02	1	1
2	2	0.27	0.12	0.05	0.02	3	2
3	2	0.28	0.08	0.05	0.02	1	2
4	3	0.41	0.10	0.05	0.02	2	3
5	1	0.46	0.11	0.05	0.02	2	1
6	3	0.32	0.13	0.05	0.02	3	3
7	3	0.30	0.16	0.05	0.02	1	3
8	3	0.34	0.07	0.05	0.02	1	3
9	1	0.54	0.12	0.05	0.02	3	1
10	2	0.39	0.19	0.05	0.02	2	2

Therefore, we could compute the regression coefficients of two moderators and their variances using Equation (3.33) and (3.34). Compared with the example with a moderator in two-level model, the weight matrix here is a block matrix instead of a diagonal matrix. The coefficient of level-2 moderator is 0.103 and variance is 0.014, and the coefficient of level-3 moderator is 0.069 and variance is 0.013 in the example. The results could be written as

$$\begin{bmatrix} \hat{\beta}_1^{\dagger} \\ \hat{\gamma}_1^{\dagger} \end{bmatrix} = \begin{bmatrix} 0.103 \\ 0.069 \end{bmatrix} \; and \; \begin{bmatrix} Var(\hat{\beta}_1^{\dagger}) \\ Var(\hat{\gamma}_1^{\dagger}) \end{bmatrix} = \begin{bmatrix} 0.014 \\ 0.013 \end{bmatrix}.$$

Then, we follow Equation (3.11) to get the non-centrality parameters (for level-2 moderator is $\lambda_{2,1}^{\dagger}$ and for level-3 moderator is $\lambda_{3,1}^{\dagger}$) of z test. The non-centrality parameters are computed as

$$\lambda_{2,1}^{\dagger} = \frac{0.103}{\sqrt{0.014}} = 0.87 \; and \; \lambda_{3,1}^{\dagger} = \frac{0.069}{\sqrt{0.013}} = 0.61.$$

The non-centrality parameters 0.87 and 0.61 in the illustrated example.

Then, we put this number into Equation (3.36) to compute the power in the case:

$$p_{2,1} = 1 - \Phi(1.96 - 0.87) + \Phi(-1.96 - 0.87) = 0.14,$$

$$p_{3,1} = 1 - \Phi(1.96 - 0.61) + \Phi(-1.96 - 0.61) = 0.09.$$

The power of the level-2 moderator in the example is 0.14 in a two-tailed Z test and the power of the level-3 moderator in the example is 0.09 in a two-tailed Z test. In the example, we see two non-significant moderators with low regression coefficients. The small non-centrality parameters indicate low power of the moderators in the three-level meta-regression model.

Moderators with Random Effects

An extension case shown in the section is when some slopes at study level are assumed random at group level. In other words, the random effects of study-level moderators exist in the model. The structure of the variance-covariance matrix is different from the previous one in Equation (3.21). To simplify the case, the present stud assumes only one moderator at the second level (study level), named X_1, is random at the third level. And only one moderator, named Z_1, is at the third level (group level). Thus, we have one more equation which represents the random effect of the slope of X_1 (which is β_{1g}). By following the three-level model structure in the chapter, the third level has two equations as

$$\beta_{0g} = \gamma_{00} + \gamma_{01} z_{1g} + \nu_{0g}, \tag{3.38}$$

$$\beta_{1g} = \gamma_{10} + \gamma_{11} z_{1g} + \nu_{1g}, \tag{3.39}$$

where $\begin{bmatrix} \nu_{0g} \\ \nu_{1g} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau^2_{(3,0)} & \tau_{(3,01)} \\ \tau_{(3,01)} & \tau^2_{(3,1)} \end{bmatrix} \right),$

Equation (3.38) shows the model for the intercept and Equation (3.39) shows the model for the level-2 slope. The two errors follow a joint distribution with means equal to 0 and variances shown in a two-by-two matrix.

In a single notation, the effect size could be written as

$$T_{ig} = \gamma_{00}+\gamma_{01}W_1 + (\gamma_{10} + \gamma_{11}W_1)X_{i1g} + \nu_{0g} + \nu_{1g}X_{i1g} + \eta_{ig} + \varepsilon_{ig}. \tag{3.40}$$

The error relates to the predictor (moderator) at study level. Equation (3.41) and (3.42) could be used to construct the variance-covariance matrix in each group. A block matrix is constructed finally to include all matrices for groups on the diagonal.

The variance for study in group g can be expressed as:

$$Variance(T_{ig}) = \left(v_i + \tau_{(2)}^2 + \tau_{(3,0)}^2\right) + 2\tau_{(3,01)}X_{i1g} + \tau_{(3,1)}^2X_{i1g}^2. \tag{3.41}$$

The covariance between studies in same group g can be expressed as:

$$Covariance(T_{ig}, T_{jg}) = \tau_{(3,0)}^2 + \tau_{(3,01)}(X_{i1g} + X_{j1g}) + \tau_{(3,1)}^2X_{i1g}X_{j1g}. \tag{3.42}$$

The following produces to compute the non-centrality parameters and the power are same to the previous section. First, we need to construct the variance-covariance matrix and invert it to get the weight matrix. Second, the weight matrix is used to compute regression coefficients of moderators. Third, the non-centrality parameters could be computed by regression coefficients and their variances. Finally, we can get the power statistics. It is important to know that adding more random effects of the moderators will lead to more complicated components in the variance-covariance matrix of error term.

Because the present scenario is more complex than before, to illustrate a simple case, an example with a smaller sample size is shown here. The case only uses two groups from previous examples. Table 3.5 shows the parameters which are needed in the following computations.

Table 3.5 An illustrated three-level meta-analysis sample with moderators and random slope

Study ID	Group ID	Effect size (T_{ig})	Within-study variance (v_i)	Between-study variance ($\tau^2_{(2)}$)	Moderator level-2 (x_{1i})
1	1	0.42	0.13	0.05	1
2	2	0.27	0.12	0.05	3
3	2	0.28	0.08	0.05	1
4	1	0.46	0.11	0.05	2
5	1	0.54	0.12	0.05	3
6	2	0.39	0.19	0.05	2
Study ID	**Group ID**	**Intercept variance ($\tau^2_{(3,0)}$)**	**Covariance ($\tau^2_{(3,01)}$)**	**Slope variance ($\tau^2_{(3,1)}$)**	**Moderator level-3 (z_1)**
1	1	0.02	0.01	0.02	1
2	2	0.02	0.01	0.02	2
3	2	0.02	0.01	0.02	2
4	1	0.02	0.01	0.02	1
5	1	0.02	0.01	0.02	1
6	2	0.02	0.01	0.02	2

The variance-covariance matrix ($\boldsymbol{V}_3^*$) could be constructed for each group by Equation (3.41) and Equation (3.42) using variance and covariances components in Table 3.5. And the inverse of the matrix is the weigh matrix ($\boldsymbol{W}_3^*$) in the case. All numbers are round to two decimals. The matrix is shown as

$$\boldsymbol{W}_3^* = (\boldsymbol{V}_3^*)^{-1} = \begin{bmatrix} \begin{bmatrix} 0.22 & 0.03 & 0.04 \\ 0.03 & 0.22 & 0.05 \\ 0.04 & 0.05 & 0.25 \end{bmatrix} & \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} \\ \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} & \begin{bmatrix} 0.25 & 0.04 & 0.05 \\ 0.04 & 0.17 & 0.03 \\ 0.05 & 0.03 & 0.30 \end{bmatrix} \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} \begin{bmatrix} 4.74 & -0.51 & -0.67 \\ -0.51 & 4.80 & -0.92 \\ -0.67 & -0.92 & 4.24 \end{bmatrix} & \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} \\ \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} & \begin{bmatrix} 4.23 & -0.91 & -0.65 \\ -0.91 & 6.18 & -0.48 \\ -0.65 & -0.48 & 3.48 \end{bmatrix} \end{bmatrix}$$

Furthermore, we could compute the regression coefficients of two moderators and their variances using Equation (3.33) and (3.34). The coefficient of level-2 moderator with a random effect is 0.102 and variance is 0.033, and the coefficient of level-3 moderator is 0.102 and variance is 0.060 in the example. The matrices of the estimated coefficients and their variance are

$$\begin{bmatrix} \hat{\beta}_1^{\dagger} \\ \hat{\gamma}_1^{\dagger} \end{bmatrix} = \begin{bmatrix} 0.102 \\ 0.102 \end{bmatrix} \; and \; \begin{bmatrix} Var(\hat{\beta}_1^{\dagger}) \\ Var(\hat{\gamma}_1^{\dagger}) \end{bmatrix} = \begin{bmatrix} 0.033 \\ 0.060 \end{bmatrix}.$$

Then, we follow Equation (3.11) to get the non-centrality parameters (for level-2 moderator is $\lambda_{2,1}^{\dagger}$ and for level-3 moderator is $\lambda_{3,1}^{\dagger}$). The non-centrality parameters are

$$\lambda_{2,1}^{\dagger} = \frac{0.102}{\sqrt{0.033}} = 0.56 \; and \; \lambda_{3,1}^{\dagger} = \frac{0.102}{\sqrt{0.060}} = 0.42.$$

Then, we put this number into Equation (3.36) to compute the power in the case. The power of the level-2 moderator in the example is 0.09 in a two-tailed Z test and the power of the level-3 moderator in the example is 0.07 in a two-tailed Z test. The negligible power statistics are explainable because the sample size is small and multiple parts of variances/covariance are assumed in the example. Thus, the regression coefficients of moderators are small, and their standard error are large. The non-significant moderators have low power values. To note that, the illustrated examples only show the ways to find non-centrality parameters and compute power statistics. All values are assumed in the examples. The computation equations are shown as

$$p_{2,1} = 1 - \Phi(1.96 - 0.56) + \Phi(-1.96 - 0.56) = 0.09,$$

$$p_{3,1} = 1 - \Phi(1.96 - 0.42) + \Phi(-1.96 - 0.42) = 0.07.$$

Simulation Study

The simulation examples in the last part aimed to show how different ratios of between-group variation in total variation affect the non-centrality parameter of the z-test and ultimately power of the weighted average effect size. A three-level meta-regression model has three parts of error variance from different levels. Two intraclass correlations (ICC) represent the relationships among three variance components, which are defined as

$$\rho_{(2)} = \frac{\tau_{(2)}^2}{\tau_{(3)}^2 + \tau_{(2)}^2 + v}, \tag{3.43}$$

$$\rho_{(3)} = \frac{\tau_{(3)}^2}{\tau_{(3)}^2 + \tau_{(2)}^2 + v}, \tag{3.44}$$

where $\rho_{(2)}$ represents the proportion of between-study variance in the total variance and $\rho_{(3)}$ represents the proportion of between-group variance in the total variance. The sum of two ICCs indicates the variances from higher levels (level-2 and level-3). To simplify the case and present the main idea of the simulation, the present study followed Hedges and Pigott's (2001) procedure, taking all sampling variance v_i to an equal value v approximately. From Equation (3.45) and (3.46), we can get the value of variance component $\tau_{(2)}^2$ and $\tau_{(3)}^2$ if we know two ICC values and the sampling variance v. The variance components could be expressed as:

$$\tau_{(2)}^2 = \frac{\rho_{(2)}}{1 - \rho_{(2)} - \rho_{(3)}} v, \tag{3.45}$$

$$\tau_{(3)}^2 = \frac{\rho_{(3)}}{1 - \rho_{(2)} - \rho_{(3)}} v. \tag{3.46}$$

Design

The simulation study assumed two population effect sizes, a moderate effect size 0.4 and a small effect size 0.2. The values of effect sizes 0.2 and 0.4 could show the variations of power statistics with different combinations of the values of variance components. Too large effect sizes lead to minor variation of the power and too small effect sizes would cause very low power in the meta-analysis. The number of studies in each group was from 2 to 10. The number of groups was assumed to 6 and 10. Thus, the range of total number of studies in the meta-regression model was from 12 to 100. The range of sample size covered usual sample sizes in empirical meta-analysis studies. The range of error variance was set to 0.05 to 0.3, which indicates a range from a small variance to a large variance. The range of ICC value was from 0.05 to 0.30 at level-2 and level-3. The sum of two ICCs covered the values from a small amount of heterogeneity to a large amount of heterogeneity (0.01 to 0.6). Overall, the design numbers of parameters are summarized in Table 3.6. The simulation study used a balanced case to illustrate the results, which means each group has the same number of studies. A two-tailed Z test was used to calculate the power statistics. For each power analysis with different parameters, the study did 1000 times iteration and finally took an average of the power statistics to control bias from randomly sampling and get a robust result. An example code is appended in Appendix C.

Table 3.6 Design numbers in simulation

Population effect size	0.2	0.4				
Numbers of group (N.group)	6	10				
Numbers of study per group (N.study)	2	4	6	8	10	
Sampling variance (within-study)	0.05	0.1	0.2	0.3		
Level-2 ICC	0.05	0.1	0.15	0.2	0.25	0.3
Level-3 ICC	0.05	0.1	0.15	0.2	0.25	0.3

Results

The results firstly displayed the power statistics from models with a medium population effect size (0.4). Four tables (Table 3.7 to Table 3.10) show average power statistics (taking average after 1000 iterations) with different parameters when the models have a different within-study variance. All numbers were round to two decimals in the tables. The results in Table 3.7 showed that the values of power were almost one due to a small within-study variance ($v = 0.05$). The change of level-2 and level-3 ICC did not strongly influence the power statistics. Although the case with the smallest sample size from the simulation (six groups and two studies per group) presented smaller values of power than other cases, all values were larger than 0.8, indicating a good power of the weighted average effect size in meta-analysis. When the numbers of study per group increased, the power increased, and when the numbers of group increased, the power increased.

Table 3.8 shows the results when the sampling variance (within-study) becomes 0.1 and the population effect size is still 0.4. The cases with ten groups had good power statistics even with considerable heterogeneity at higher levels. Larger sample sizes and smaller level-2/level-3 ICC gave higher statistical powers (near to one). However, the cases with six groups and two studies in each group, locating at the first block in the table, showed some powers were lower than 0.8. For instance, when level-2 ICC and level-3 ICC were higher than 0.2, the power values were all smaller than 0.8. It indicated that when the proportion of variance at higher levels became larger, the power went lower. These changes were more visible than the changes in the models with a minor sampling variance (0.05).

Table 3.9 shows the results when the within-study variance becomes 0.2 and the population effect size is still 0.4. The values of power went lower than those in Table 3.7 because of the larger within-study sampling variance. When level-3 ICC was 0.3, which indicated 30% of the variance came from the group level, the power was smaller than the cases with smaller level-3 ICC. Especially when the number of groups was small (e.g., six groups), almost all values of power were lower than 0.80. If the number of studies in each group is small, the power statistics went down to 0.5. However, the simulation results also showed that if a meta-analysis has a good sample size, for instance, 60 studies or more, the analysis using a three-level model (considering between-group variance) could still have a good power of weighted average effect size.

Table 3.10 shows the results when the within-study variance increased to 0.3 and the population effect size was still 0.4. The simulation examples with high level-3 ICC (e.g., 0.25, 0.3) displayed low power statistics even the meta-analysis had a considerable sample size. Compared with the results from the previous tables, the results exposed that large within-study variance could influence the power strongly. Also, large level-2 and level-3 ICC influenced the power statistics strongly. When the case had the same level-2 ICC, larger level-3 ICC impacted the power significantly, especially for a small sample size case. In conclusion, the results indicated that researchers need to pay more attention when the meta-analysis has a large within-study variance (larger than 0.1).

Table 3.7 Power in the models with population effect size 0.4 and within-study variance 0.05

		N.group = 6						N.group = 10					
		ICC3						ICC3					
N.study = 2		**0.005**	**0.1**	**0.15**	**0.2**	**0.25**	**0.3**	**0.005**	**0.1**	**0.15**	**0.2**	**0.25**	**0.3**
	0.005	1.00	0.99	0.99	0.98	0.98	0.97	1.00	1.00	1.00	1.00	1.00	1.00
	0.1	1.00	0.99	0.98	0.97	0.97	0.95	1.00	1.00	1.00	1.00	1.00	0.99
ICC2	**0.15**	0.99	0.99	0.98	0.97	0.95	0.93	1.00	1.00	1.00	1.00	1.00	0.99
	0.2	0.99	0.98	0.97	0.96	0.94	0.92	1.00	1.00	1.00	1.00	0.99	0.99
	0.25	0.99	0.97	0.96	0.94	0.91	0.88	1.00	1.00	1.00	0.99	0.98	0.97
	0.3	0.99	0.97	0.95	0.93	0.90	0.87	1.00	1.00	0.99	0.99	0.98	0.96
N.study = 4													
	0.005	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00
	0.1	1.00	1.00	1.00	1.00	0.99	0.98	1.00	1.00	1.00	1.00	1.00	1.00
ICC2	**0.15**	1.00	1.00	1.00	1.00	0.99	0.98	1.00	1.00	1.00	1.00	1.00	1.00
	0.2	1.00	1.00	1.00	0.99	0.98	0.96	1.00	1.00	1.00	1.00	1.00	1.00
	0.25	1.00	1.00	1.00	0.99	0.97	0.95	1.00	1.00	1.00	1.00	1.00	0.99
	0.3	1.00	1.00	0.99	0.98	0.96	0.93	1.00	1.00	1.00	1.00	1.00	0.99
N.study = 6													
	0.005	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.1	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00
ICC2	**0.15**	1.00	1.00	1.00	1.00	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00
	0.2	1.00	1.00	1.00	1.00	0.99	0.98	1.00	1.00	1.00	1.00	1.00	1.00
	0.25	1.00	1.00	1.00	0.99	0.98	0.97	1.00	1.00	1.00	1.00	1.00	1.00
	0.3	1.00	1.00	1.00	0.99	0.98	0.95	1.00	1.00	1.00	1.00	1.00	0.99
N.study = 8													
	0.005	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.1	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00
ICC2	**0.15**	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00
	0.2	1.00	1.00	1.00	1.00	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00
	0.25	1.00	1.00	1.00	1.00	0.99	0.97	1.00	1.00	1.00	1.00	1.00	1.00
	0.3	1.00	1.00	1.00	1.00	0.98	0.96	1.00	1.00	1.00	1.00	1.00	1.00
N.study = 10													
	0.005	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.1	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00
ICC2	**0.15**	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00
	0.2	1.00	1.00	1.00	1.00	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00
	0.25	1.00	1.00	1.00	1.00	0.99	0.98	1.00	1.00	1.00	1.00	1.00	1.00
	0.3	1.00	1.00	1.00	1.00	0.99	0.97	1.00	1.00	1.00	1.00	1.00	1.00

Note. The powers are calculated based on two-tailed Z tests.

Table 3.8 Power in the models with population effect size 0.4 and within-study variance 0.1

		N.group = 6						N.group = 10					
		ICC3						ICC3					
N.study = 2		**0.005**	**0.1**	**0.15**	**0.2**	**0.25**	**0.3**	**0.005**	**0.1**	**0.15**	**0.2**	**0.25**	**0.3**
	0.005	0.96	0.92	0.90	0.86	0.84	0.82	1.00	0.99	0.98	0.97	0.95	0.94
	0.1	0.93	0.89	0.86	0.84	0.81	0.78	0.99	0.98	0.97	0.95	0.94	0.91
ICC2	**0.15**	0.92	0.88	0.85	0.81	0.77	0.74	0.99	0.97	0.96	0.94	0.92	0.90
	0.2	0.90	0.86	0.82	0.80	0.75	0.72	0.99	0.97	0.95	0.93	0.90	0.87
	0.25	0.90	0.83	0.80	0.75	0.71	0.66	0.98	0.96	0.93	0.90	0.86	0.83
	0.3	0.88	0.81	0.77	0.74	0.68	0.65	0.97	0.95	0.91	0.88	0.85	0.80
N.study = 4													
	0.005	1.00	0.99	0.98	0.95	0.92	0.90	1.00	1.00	1.00	0.99	0.99	0.98
	0.1	1.00	0.98	0.96	0.93	0.90	0.86	1.00	1.00	0.99	0.99	0.98	0.97
ICC2	**0.15**	0.99	0.97	0.96	0.92	0.89	0.85	1.00	1.00	0.99	0.99	0.97	0.95
	0.2	0.99	0.97	0.94	0.90	0.85	0.80	1.00	1.00	0.99	0.98	0.96	0.94
	0.25	0.99	0.95	0.93	0.88	0.83	0.78	1.00	1.00	0.99	0.97	0.95	0.91
	0.3	0.99	0.95	0.90	0.85	0.79	0.73	1.00	0.99	0.98	0.96	0.93	0.89
N.study = 6													
	0.005	1.00	1.00	0.99	0.98	0.96	0.92	1.00	1.00	1.00	1.00	1.00	0.99
	0.1	1.00	0.99	0.98	0.96	0.93	0.89	1.00	1.00	1.00	1.00	0.99	0.98
ICC2	**0.15**	1.00	0.99	0.97	0.95	0.92	0.88	1.00	1.00	1.00	1.00	0.99	0.97
	0.2	1.00	0.99	0.97	0.94	0.91	0.86	1.00	1.00	1.00	0.99	0.98	0.96
	0.25	1.00	0.99	0.97	0.92	0.86	0.82	1.00	1.00	1.00	0.99	0.97	0.94
	0.3	1.00	0.98	0.94	0.90	0.85	0.78	1.00	1.00	0.99	0.98	0.95	0.91
N.study = 8													
	0.005	1.00	1.00	0.99	0.99	0.96	0.94	1.00	1.00	1.00	1.00	1.00	0.99
	0.1	1.00	1.00	0.99	0.98	0.95	0.91	1.00	1.00	1.00	1.00	0.99	0.98
ICC2	**0.15**	1.00	1.00	0.99	0.97	0.94	0.89	1.00	1.00	1.00	1.00	0.99	0.98
	0.2	1.00	0.99	0.98	0.96	0.91	0.87	1.00	1.00	1.00	1.00	0.99	0.97
	0.25	1.00	0.99	0.97	0.95	0.90	0.83	1.00	1.00	1.00	0.99	0.98	0.95
	0.3	1.00	0.99	0.97	0.92	0.87	0.80	1.00	1.00	1.00	0.99	0.97	0.93
N.study = 10													
	0.005	1.00	1.00	1.00	0.99	0.97	0.95	1.00	1.00	1.00	1.00	1.00	0.99
	0.1	1.00	1.00	0.99	0.98	0.96	0.91	1.00	1.00	1.00	1.00	1.00	0.99
ICC2	**0.15**	1.00	1.00	0.99	0.97	0.94	0.90	1.00	1.00	1.00	1.00	0.99	0.98
	0.2	1.00	1.00	0.99	0.97	0.92	0.88	1.00	1.00	1.00	1.00	0.99	0.97
	0.25	1.00	1.00	0.98	0.95	0.91	0.84	1.00	1.00	1.00	0.99	0.98	0.96
	0.3	1.00	0.99	0.97	0.94	0.88	0.81	1.00	1.00	1.00	0.99	0.97	0.94

Note. The powers are calculated based on two-tailed Z tests.

Table 3.9 Power in the models with population effect size 0.4 and within-study variance 0.2

		N.group = 6						**N.group = 10**					
		ICC3						**ICC3**					
N.study = 2		**0.005**	**0.1**	**0.15**	**0.2**	**0.25**	**0.3**	**0.005**	**0.1**	**0.15**	**0.2**	**0.25**	**0.3**
	0.005	0.79	0.72	0.69	0.65	0.62	0.60	0.93	0.88	0.85	0.82	0.77	0.76
	0.1	0.75	0.68	0.65	0.62	0.59	0.56	0.90	0.85	0.81	0.77	0.74	0.71
ICC2	**0.15**	0.73	0.66	0.62	0.58	0.54	0.52	0.88	0.83	0.79	0.75	0.72	0.69
	0.2	0.70	0.64	0.59	0.58	0.53	0.50	0.87	0.82	0.78	0.74	0.69	0.65
	0.25	0.69	0.60	0.57	0.53	0.49	0.46	0.86	0.79	0.73	0.69	0.64	0.61
	0.3	0.66	0.58	0.55	0.52	0.47	0.46	0.84	0.77	0.71	0.67	0.62	0.57
N.study = 4													
	0.005	0.95	0.89	0.84	0.79	0.73	0.70	1.00	0.97	0.95	0.92	0.89	0.85
	0.1	0.94	0.85	0.80	0.74	0.70	0.65	0.99	0.96	0.93	0.89	0.86	0.82
ICC2	**0.15**	0.92	0.83	0.78	0.72	0.68	0.63	0.99	0.95	0.91	0.88	0.83	0.77
	0.2	0.91	0.81	0.76	0.70	0.63	0.57	0.99	0.94	0.90	0.85	0.80	0.76
	0.25	0.90	0.78	0.73	0.66	0.60	0.55	0.98	0.93	0.88	0.83	0.78	0.71
	0.3	0.87	0.77	0.70	0.63	0.56	0.51	0.97	0.91	0.86	0.81	0.74	0.67
N.study = 6													
	0.005	0.99	0.94	0.90	0.84	0.79	0.73	1.00	0.99	0.98	0.95	0.93	0.87
	0.1	0.99	0.91	0.87	0.81	0.74	0.68	1.00	0.98	0.96	0.93	0.90	0.85
ICC2	**0.15**	0.98	0.90	0.84	0.78	0.72	0.67	1.00	0.98	0.95	0.92	0.88	0.82
	0.2	0.98	0.88	0.82	0.75	0.71	0.64	1.00	0.98	0.94	0.91	0.86	0.79
	0.25	0.96	0.87	0.81	0.73	0.64	0.59	1.00	0.97	0.94	0.88	0.82	0.75
	0.3	0.96	0.85	0.77	0.70	0.63	0.56	1.00	0.96	0.92	0.86	0.78	0.72
N.study = 8													
	0.005	1.00	0.97	0.92	0.88	0.80	0.76	1.00	1.00	0.99	0.97	0.94	0.90
	0.1	1.00	0.94	0.90	0.85	0.78	0.70	1.00	0.99	0.98	0.95	0.92	0.87
ICC2	**0.15**	0.99	0.94	0.88	0.81	0.75	0.69	1.00	0.99	0.97	0.94	0.89	0.85
	0.2	0.99	0.92	0.86	0.80	0.70	0.66	1.00	0.99	0.97	0.93	0.87	0.81
	0.25	0.99	0.91	0.84	0.77	0.70	0.60	1.00	0.98	0.95	0.90	0.85	0.78
	0.3	0.99	0.88	0.81	0.72	0.65	0.57	1.00	0.98	0.94	0.89	0.82	0.73
N.study = 10													
	0.005	1.00	0.98	0.94	0.89	0.82	0.77	1.00	1.00	0.99	0.98	0.95	0.91
	0.1	1.00	0.96	0.91	0.86	0.79	0.71	1.00	1.00	0.98	0.96	0.92	0.88
ICC2	**0.15**	1.00	0.95	0.90	0.84	0.76	0.70	1.00	1.00	0.98	0.95	0.91	0.85
	0.2	1.00	0.95	0.89	0.81	0.73	0.66	1.00	0.99	0.97	0.94	0.89	0.83
	0.25	1.00	0.93	0.86	0.78	0.71	0.62	1.00	0.99	0.97	0.91	0.86	0.80
	0.3	1.00	0.92	0.84	0.76	0.68	0.59	1.00	0.98	0.96	0.90	0.83	0.74

Note. The powers are calculated based on two-tailed Z tests.

Table 3.10 Power in the models with population effect size 0.4 and within-study variance 0.3

		N.group = 6						N.group = 10					
		ICC3						ICC3					
N.study = 2		**0.005**	**0.1**	**0.15**	**0.2**	**0.25**	**0.3**	**0.005**	**0.1**	**0.15**	**0.2**	**0.25**	**0.3**
	0.005	0.65	0.59	0.57	0.52	0.50	0.48	0.83	0.76	0.72	0.69	0.64	0.63
	0.1	0.62	0.55	0.53	0.50	0.48	0.45	0.78	0.73	0.68	0.64	0.61	0.58
ICC2	**0.15**	0.60	0.54	0.50	0.47	0.43	0.42	0.76	0.70	0.66	0.62	0.58	0.56
	0.2	0.57	0.52	0.47	0.47	0.43	0.40	0.75	0.69	0.64	0.61	0.57	0.52
	0.25	0.56	0.48	0.46	0.43	0.40	0.37	0.73	0.65	0.60	0.56	0.51	0.49
	0.3	0.53	0.47	0.44	0.42	0.38	0.37	0.71	0.64	0.58	0.54	0.50	0.46
N.study = 4													
	0.005	0.86	0.77	0.71	0.66	0.60	0.57	0.97	0.91	0.87	0.82	0.78	0.73
	0.1	0.84	0.72	0.66	0.61	0.58	0.53	0.95	0.88	0.82	0.78	0.74	0.68
ICC2	**0.15**	0.82	0.70	0.65	0.58	0.55	0.51	0.94	0.86	0.81	0.76	0.70	0.63
	0.2	0.80	0.68	0.62	0.57	0.51	0.46	0.94	0.85	0.79	0.72	0.67	0.63
	0.25	0.79	0.65	0.59	0.54	0.48	0.44	0.93	0.82	0.76	0.70	0.64	0.58
	0.3	0.75	0.63	0.57	0.50	0.45	0.41	0.91	0.80	0.74	0.67	0.60	0.54
N.study = 6													
	0.005	0.95	0.84	0.78	0.72	0.66	0.60	1.00	0.95	0.91	0.87	0.82	0.75
	0.1	0.94	0.80	0.75	0.68	0.61	0.55	0.99	0.93	0.89	0.83	0.78	0.73
ICC2	**0.15**	0.92	0.79	0.71	0.65	0.58	0.55	0.98	0.93	0.87	0.82	0.76	0.69
	0.2	0.92	0.77	0.69	0.61	0.58	0.51	0.98	0.92	0.85	0.80	0.74	0.65
	0.25	0.89	0.74	0.68	0.60	0.52	0.47	0.98	0.90	0.84	0.76	0.69	0.61
	0.3	0.89	0.72	0.63	0.57	0.51	0.45	0.97	0.87	0.82	0.73	0.65	0.59
N.study = 8													
	0.005	0.98	0.89	0.81	0.76	0.67	0.63	1.00	0.97	0.94	0.91	0.84	0.78
	0.1	0.98	0.85	0.79	0.72	0.64	0.57	1.00	0.96	0.92	0.86	0.80	0.75
ICC2	**0.15**	0.97	0.84	0.75	0.67	0.62	0.56	1.00	0.95	0.91	0.85	0.78	0.72
	0.2	0.96	0.81	0.73	0.66	0.57	0.53	1.00	0.94	0.89	0.83	0.75	0.68
	0.25	0.95	0.79	0.71	0.64	0.58	0.48	0.99	0.92	0.87	0.79	0.72	0.65
	0.3	0.94	0.76	0.67	0.59	0.52	0.46	0.99	0.91	0.85	0.77	0.69	0.60
N.study = 10													
	0.005	1.00	0.91	0.84	0.78	0.69	0.64	1.00	0.98	0.96	0.91	0.87	0.80
	0.1	0.99	0.89	0.79	0.74	0.66	0.58	1.00	0.98	0.94	0.88	0.82	0.76
ICC2	**0.15**	0.99	0.87	0.78	0.71	0.62	0.57	1.00	0.97	0.93	0.86	0.81	0.73
	0.2	0.98	0.86	0.77	0.68	0.59	0.53	1.00	0.96	0.91	0.85	0.77	0.70
	0.25	0.98	0.82	0.74	0.65	0.58	0.50	1.00	0.95	0.90	0.80	0.74	0.67
	0.3	0.97	0.82	0.71	0.63	0.55	0.48	1.00	0.93	0.87	0.79	0.70	0.61

Note. The powers are calculated based on two-tailed Z tests.

Table 3.11 to Table 3.14 show the results when the population effect size is small (0.2).

Table 3.11 Power in the models with population effect size 0.2 and within-study variance 0.05

		N.group = 6						N.group = 10					
		ICC3						ICC3					
N.study = 2		**0.005**	**0.1**	**0.15**	**0.2**	**0.25**	**0.3**	**0.005**	**0.1**	**0.15**	**0.2**	**0.25**	**0.3**
	0.005	0.79	0.72	0.69	0.65	0.62	0.60	0.93	0.88	0.85	0.82	0.77	0.76
	0.1	0.75	0.68	0.65	0.62	0.59	0.56	0.90	0.85	0.81	0.77	0.74	0.71
ICC2	**0.15**	0.73	0.66	0.62	0.58	0.54	0.52	0.88	0.83	0.79	0.75	0.72	0.69
	0.2	0.70	0.64	0.59	0.58	0.53	0.50	0.87	0.82	0.78	0.74	0.69	0.65
	0.25	0.69	0.60	0.57	0.53	0.49	0.46	0.86	0.79	0.73	0.69	0.64	0.61
	0.3	0.66	0.58	0.55	0.52	0.47	0.46	0.84	0.77	0.71	0.67	0.62	0.57
N.study = 4													
	0.005	0.95	0.89	0.84	0.79	0.73	0.70	1.00	0.97	0.95	0.92	0.89	0.85
	0.1	0.94	0.85	0.80	0.74	0.70	0.65	0.99	0.96	0.93	0.89	0.86	0.82
ICC2	**0.15**	0.92	0.83	0.78	0.72	0.68	0.63	0.99	0.95	0.91	0.88	0.83	0.77
	0.2	0.91	0.81	0.76	0.70	0.63	0.57	0.99	0.94	0.90	0.85	0.80	0.76
	0.25	0.90	0.78	0.73	0.66	0.60	0.55	0.98	0.93	0.88	0.83	0.78	0.71
	0.3	0.87	0.77	0.70	0.63	0.56	0.51	0.97	0.91	0.86	0.81	0.74	0.67
N.study = 6													
	0.005	0.99	0.94	0.90	0.84	0.79	0.73	1.00	0.99	0.98	0.95	0.93	0.87
	0.1	0.99	0.91	0.87	0.81	0.74	0.68	1.00	0.98	0.96	0.93	0.90	0.85
ICC2	**0.15**	0.98	0.90	0.84	0.78	0.72	0.67	1.00	0.98	0.95	0.92	0.88	0.82
	0.2	0.98	0.88	0.82	0.75	0.71	0.64	1.00	0.98	0.94	0.91	0.86	0.79
	0.25	0.96	0.87	0.81	0.73	0.64	0.59	1.00	0.97	0.94	0.88	0.82	0.75
	0.3	0.96	0.85	0.77	0.70	0.63	0.56	1.00	0.96	0.92	0.86	0.78	0.72
N.study = 8													
	0.005	1.00	0.97	0.92	0.88	0.80	0.76	1.00	1.00	0.99	0.97	0.94	0.90
	0.1	1.00	0.94	0.90	0.85	0.78	0.70	1.00	0.99	0.98	0.95	0.92	0.87
ICC2	**0.15**	0.99	0.94	0.88	0.81	0.75	0.69	1.00	0.99	0.97	0.94	0.89	0.85
	0.2	0.99	0.92	0.86	0.80	0.70	0.66	1.00	0.99	0.97	0.93	0.87	0.81
	0.25	0.99	0.91	0.84	0.77	0.70	0.60	1.00	0.98	0.95	0.90	0.85	0.78
	0.3	0.99	0.88	0.81	0.72	0.65	0.57	1.00	0.98	0.94	0.89	0.82	0.73
N.study = 10													
	0.005	1.00	0.98	0.94	0.89	0.82	0.77	1.00	1.00	0.99	0.98	0.95	0.91
	0.1	1.00	0.96	0.91	0.86	0.79	0.71	1.00	1.00	0.98	0.96	0.92	0.88
ICC2	**0.15**	1.00	0.95	0.90	0.84	0.76	0.70	1.00	1.00	0.98	0.95	0.91	0.85
	0.2	1.00	0.95	0.89	0.81	0.73	0.66	1.00	0.99	0.97	0.94	0.89	0.83
	0.25	1.00	0.93	0.86	0.78	0.71	0.62	1.00	0.99	0.97	0.91	0.86	0.80
	0.3	1.00	0.92	0.84	0.76	0.68	0.59	1.00	0.98	0.96	0.90	0.83	0.74

Note. The powers are calculated based on two-tailed Z tests.

Table 3.12 Power in the models with population effect size 0.2 and within-study variance 0.1

		N.group = 6						N.group = 10					
		ICC3						ICC3					
N.study = 2		**0.005**	**0.1**	**0.15**	**0.2**	**0.25**	**0.3**	**0.005**	**0.1**	**0.15**	**0.2**	**0.25**	**0.3**
	0.005	0.56	0.51	0.49	0.45	0.43	0.42	0.73	0.66	0.65	0.58	0.58	0.52
	0.1	0.53	0.47	0.45	0.43	0.41	0.39	0.70	0.62	0.59	0.57	0.53	0.49
ICC2	**0.15**	0.52	0.46	0.43	0.40	0.37	0.36	0.68	0.60	0.57	0.54	0.51	0.47
	0.2	0.48	0.44	0.41	0.41	0.37	0.35	0.66	0.58	0.58	0.50	0.48	0.46
	0.25	0.48	0.41	0.39	0.37	0.34	0.32	0.63	0.55	0.52	0.50	0.44	0.42
	0.3	0.46	0.40	0.38	0.36	0.33	0.33	0.60	0.55	0.50	0.46	0.43	0.39
N.study = 4													
	0.005	0.77	0.68	0.62	0.57	0.51	0.49	0.92	0.83	0.78	0.72	0.67	0.60
	0.1	0.75	0.62	0.57	0.52	0.50	0.45	0.89	0.80	0.75	0.68	0.64	0.58
ICC2	**0.15**	0.73	0.61	0.56	0.50	0.48	0.44	0.88	0.78	0.71	0.67	0.61	0.56
	0.2	0.71	0.58	0.53	0.49	0.43	0.40	0.88	0.75	0.70	0.64	0.58	0.53
	0.25	0.70	0.56	0.51	0.46	0.42	0.38	0.86	0.73	0.66	0.60	0.58	0.51
	0.3	0.66	0.55	0.49	0.43	0.39	0.35	0.84	0.70	0.63	0.58	0.53	0.47
N.study = 6													
	0.005	0.89	0.76	0.69	0.62	0.57	0.51	0.98	0.90	0.85	0.79	0.72	0.68
	0.1	0.88	0.71	0.65	0.58	0.52	0.47	0.97	0.88	0.81	0.75	0.69	0.64
ICC2	**0.15**	0.85	0.69	0.62	0.56	0.50	0.47	0.96	0.86	0.79	0.74	0.66	0.58
	0.2	0.85	0.67	0.59	0.53	0.50	0.44	0.95	0.84	0.77	0.69	0.64	0.55
	0.25	0.81	0.64	0.59	0.52	0.44	0.40	0.95	0.82	0.74	0.67	0.60	0.53
	0.3	0.81	0.63	0.55	0.49	0.44	0.39	0.93	0.79	0.71	0.64	0.56	0.51
N.study = 8													
	0.005	0.95	0.81	0.72	0.66	0.58	0.54	0.99	0.94	0.88	0.82	0.75	0.68
	0.1	0.94	0.77	0.69	0.63	0.56	0.49	0.99	0.91	0.84	0.77	0.72	0.66
ICC2	**0.15**	0.92	0.75	0.66	0.58	0.53	0.48	0.99	0.90	0.83	0.76	0.69	0.61
	0.2	0.91	0.72	0.64	0.57	0.49	0.46	0.98	0.88	0.81	0.73	0.64	0.60
	0.25	0.89	0.69	0.62	0.55	0.50	0.41	0.98	0.86	0.77	0.70	0.63	0.54
	0.3	0.88	0.67	0.58	0.50	0.45	0.40	0.97	0.84	0.76	0.66	0.61	0.52
N.study = 10													
	0.005	0.98	0.84	0.75	0.68	0.59	0.55	1.00	0.95	0.91	0.85	0.77	0.71
	0.1	0.97	0.81	0.70	0.64	0.57	0.50	1.00	0.94	0.87	0.80	0.74	0.67
ICC2	**0.15**	0.95	0.78	0.69	0.62	0.54	0.49	1.00	0.92	0.86	0.77	0.70	0.63
	0.2	0.95	0.78	0.68	0.59	0.51	0.45	0.99	0.90	0.84	0.75	0.66	0.60
	0.25	0.94	0.73	0.64	0.57	0.50	0.42	0.99	0.89	0.79	0.73	0.65	0.56
	0.3	0.92	0.73	0.62	0.54	0.48	0.41	0.99	0.89	0.78	0.70	0.61	0.54

Note. The powers are calculated based on two-tailed Z tests.

Table 3.13 Power in the models with population effect size 0.2 and within-study variance 0.2

		N.group = 6						N.group = 10					
		ICC3						ICC3					
N.study = 2		**0.005**	**0.1**	**0.15**	**0.2**	**0.25**	**0.3**	**0.005**	**0.1**	**0.15**	**0.2**	**0.25**	**0.3**
	0.005	0.38	0.36	0.34	0.32	0.31	0.30	0.52	0.46	0.43	0.42	0.38	0.38
	0.1	0.37	0.33	0.33	0.31	0.30	0.29	0.48	0.45	0.41	0.38	0.36	0.35
ICC2	**0.15**	0.37	0.33	0.31	0.29	0.27	0.27	0.47	0.41	0.40	0.36	0.34	0.34
	0.2	0.34	0.31	0.29	0.30	0.28	0.26	0.45	0.41	0.38	0.37	0.35	0.32
	0.25	0.34	0.29	0.28	0.27	0.25	0.25	0.44	0.39	0.36	0.34	0.31	0.30
	0.3	0.32	0.29	0.28	0.27	0.24	0.25	0.43	0.38	0.35	0.33	0.30	0.28
N.study = 4													
	0.005	0.55	0.47	0.43	0.39	0.36	0.35	0.72	0.61	0.57	0.52	0.48	0.45
	0.1	0.53	0.43	0.39	0.36	0.36	0.32	0.68	0.57	0.52	0.48	0.45	0.41
ICC2	**0.15**	0.51	0.42	0.39	0.34	0.34	0.32	0.66	0.55	0.50	0.46	0.42	0.38
	0.2	0.49	0.40	0.37	0.35	0.31	0.29	0.66	0.53	0.48	0.44	0.40	0.37
	0.25	0.49	0.39	0.35	0.33	0.30	0.27	0.64	0.51	0.46	0.41	0.37	0.34
	0.3	0.46	0.38	0.34	0.30	0.28	0.26	0.61	0.50	0.46	0.40	0.35	0.32
N.study = 6													
	0.005	0.68	0.54	0.48	0.43	0.39	0.36	0.86	0.68	0.62	0.57	0.51	0.46
	0.1	0.67	0.49	0.46	0.41	0.36	0.33	0.82	0.64	0.58	0.52	0.48	0.44
ICC2	**0.15**	0.63	0.48	0.43	0.39	0.35	0.33	0.79	0.64	0.56	0.52	0.47	0.41
	0.2	0.63	0.47	0.41	0.37	0.35	0.31	0.78	0.63	0.54	0.49	0.45	0.39
	0.25	0.58	0.44	0.41	0.36	0.31	0.29	0.75	0.60	0.53	0.46	0.41	0.36
	0.3	0.58	0.44	0.38	0.34	0.31	0.28	0.74	0.56	0.52	0.44	0.38	0.35
N.study = 8													
	0.005	0.77	0.59	0.50	0.46	0.40	0.38	0.92	0.74	0.67	0.62	0.53	0.48
	0.1	0.75	0.55	0.48	0.45	0.39	0.34	0.89	0.71	0.63	0.55	0.49	0.46
ICC2	**0.15**	0.72	0.53	0.46	0.40	0.37	0.34	0.87	0.69	0.61	0.54	0.48	0.43
	0.2	0.70	0.50	0.44	0.40	0.34	0.33	0.86	0.68	0.60	0.52	0.45	0.41
	0.25	0.68	0.48	0.43	0.38	0.35	0.30	0.84	0.64	0.56	0.48	0.43	0.39
	0.3	0.66	0.46	0.40	0.35	0.31	0.29	0.83	0.62	0.54	0.48	0.42	0.35
N.study = 10													
	0.005	0.85	0.61	0.53	0.48	0.41	0.38	0.95	0.78	0.69	0.62	0.56	0.50
	0.1	0.82	0.60	0.48	0.45	0.40	0.35	0.94	0.75	0.66	0.58	0.51	0.46
ICC2	**0.15**	0.79	0.55	0.48	0.43	0.37	0.35	0.93	0.74	0.64	0.55	0.50	0.44
	0.2	0.77	0.56	0.48	0.40	0.35	0.32	0.92	0.70	0.61	0.55	0.47	0.42
	0.25	0.76	0.50	0.45	0.40	0.35	0.30	0.89	0.67	0.61	0.49	0.44	0.40
	0.3	0.72	0.51	0.43	0.38	0.34	0.30	0.89	0.65	0.57	0.49	0.42	0.36

Note. The powers are calculated based on two-tailed Z tests.

Table 3.14 Power in the models with population effect size 0.2 and within-study variance 0.3

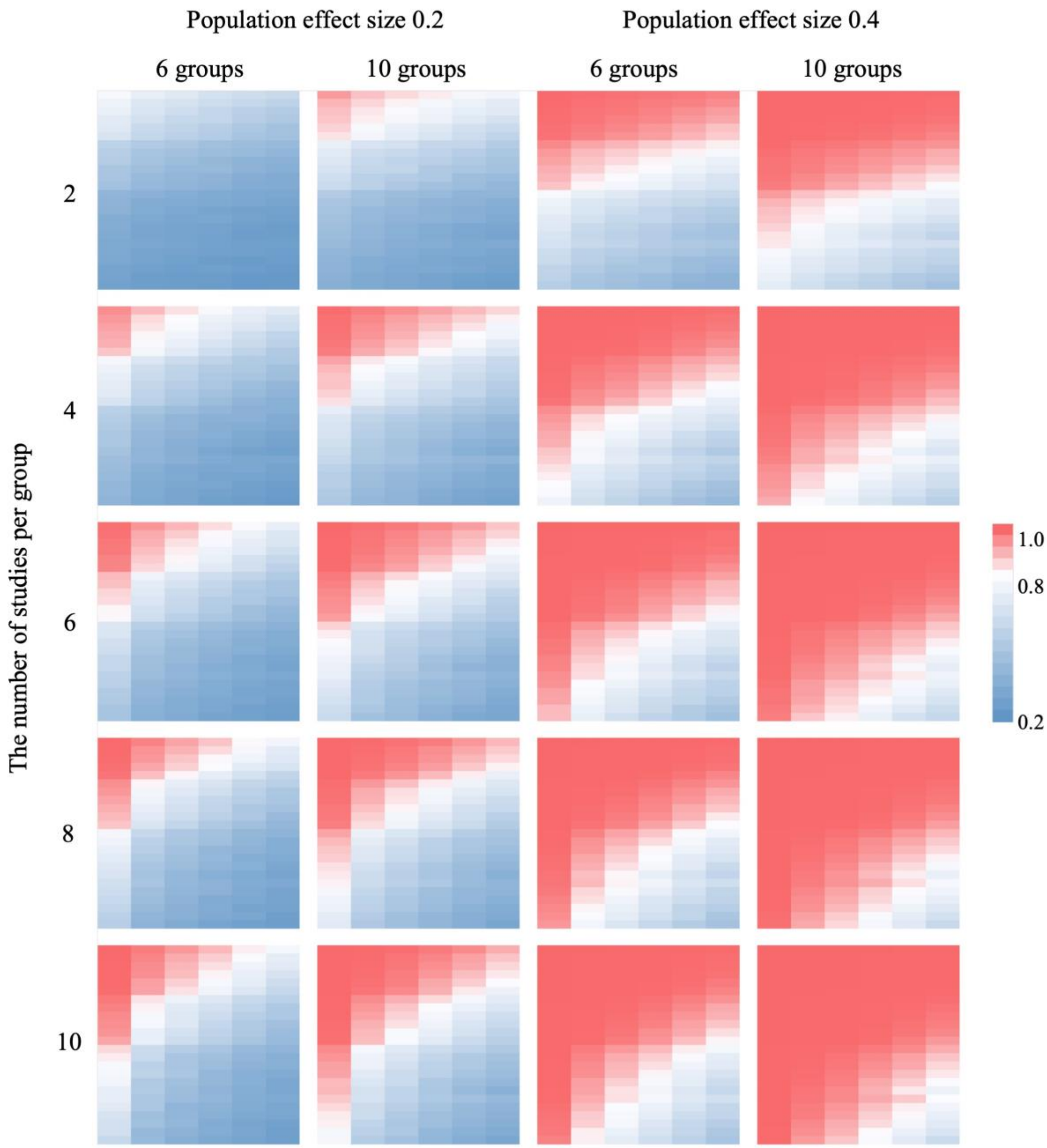
		N.group = 6						N.group = 10					
		ICC3						ICC3					
N.study = 2		**0.005**	**0.1**	**0.15**	**0.2**	**0.25**	**0.3**	**0.005**	**0.1**	**0.15**	**0.2**	**0.25**	**0.3**
	0.005	0.31	0.30	0.28	0.27	0.26	0.26	0.42	0.37	0.35	0.34	0.31	0.32
	0.1	0.31	0.28	0.28	0.26	0.26	0.25	0.39	0.37	0.34	0.31	0.30	0.29
ICC2	**0.15**	0.31	0.28	0.26	0.25	0.23	0.24	0.38	0.34	0.33	0.30	0.29	0.28
	0.2	0.28	0.26	0.25	0.26	0.24	0.23	0.37	0.34	0.31	0.31	0.29	0.27
	0.25	0.28	0.25	0.24	0.24	0.22	0.22	0.36	0.32	0.30	0.28	0.26	0.25
	0.3	0.27	0.25	0.25	0.24	0.22	0.22	0.35	0.32	0.30	0.28	0.25	0.24
N.study = 4													
	0.005	0.44	0.38	0.36	0.33	0.30	0.30	0.59	0.50	0.47	0.42	0.40	0.37
	0.1	0.43	0.35	0.31	0.30	0.30	0.27	0.55	0.46	0.42	0.39	0.37	0.33
ICC2	**0.15**	0.41	0.34	0.32	0.29	0.29	0.27	0.53	0.44	0.40	0.37	0.34	0.31
	0.2	0.40	0.33	0.30	0.29	0.26	0.25	0.53	0.43	0.39	0.36	0.33	0.31
	0.25	0.39	0.32	0.29	0.28	0.26	0.24	0.51	0.41	0.37	0.34	0.30	0.29
	0.3	0.38	0.31	0.29	0.26	0.24	0.23	0.49	0.40	0.38	0.32	0.29	0.27
N.study = 6													
	0.005	0.55	0.43	0.39	0.35	0.32	0.30	0.73	0.55	0.50	0.46	0.41	0.37
	0.1	0.55	0.40	0.37	0.33	0.30	0.28	0.69	0.52	0.46	0.42	0.39	0.36
ICC2	**0.15**	0.51	0.39	0.35	0.33	0.29	0.28	0.67	0.51	0.45	0.42	0.38	0.33
	0.2	0.51	0.38	0.33	0.31	0.29	0.26	0.64	0.51	0.44	0.40	0.37	0.32
	0.25	0.47	0.36	0.33	0.30	0.26	0.25	0.62	0.49	0.43	0.37	0.34	0.30
	0.3	0.47	0.35	0.32	0.28	0.27	0.25	0.61	0.45	0.42	0.36	0.31	0.30
N.study = 8													
	0.005	0.64	0.48	0.40	0.37	0.33	0.31	0.82	0.61	0.55	0.50	0.42	0.38
	0.1	0.62	0.44	0.39	0.37	0.32	0.28	0.77	0.58	0.51	0.44	0.39	0.38
ICC2	**0.15**	0.60	0.43	0.37	0.32	0.30	0.28	0.75	0.56	0.49	0.44	0.39	0.35
	0.2	0.58	0.41	0.36	0.33	0.28	0.28	0.74	0.55	0.48	0.42	0.37	0.33
	0.25	0.55	0.38	0.35	0.32	0.30	0.25	0.71	0.52	0.44	0.39	0.35	0.32
	0.3	0.53	0.37	0.32	0.29	0.27	0.25	0.70	0.50	0.44	0.39	0.35	0.29
N.study = 10													
	0.005	0.72	0.49	0.42	0.39	0.33	0.31	0.87	0.65	0.56	0.50	0.45	0.41
	0.1	0.69	0.49	0.38	0.36	0.33	0.29	0.85	0.61	0.54	0.47	0.41	0.38
ICC2	**0.15**	0.66	0.44	0.39	0.35	0.31	0.30	0.82	0.61	0.51	0.44	0.40	0.36
	0.2	0.64	0.45	0.39	0.33	0.29	0.27	0.82	0.57	0.49	0.44	0.38	0.34
	0.25	0.63	0.40	0.36	0.33	0.30	0.26	0.77	0.55	0.49	0.40	0.36	0.33
	0.3	0.59	0.42	0.35	0.31	0.29	0.26	0.78	0.52	0.46	0.39	0.34	0.29

Note. The powers are calculated based on two-tailed Z tests.

The simulation results from Table 3.11 to Table 3.14 indicated that the population effect size strongly influenced the power in the three-level meta-regression model. The population effect size decreased to 0.2, which indicated a small effect size. Under this scenario, the power was high only in the model with a small sampling variance (within-study) and small ICC values from higher levels. The power went lower when the within-study variance went higher from Table 3.11 to Table 3.14. Especially in Table 3.14, most power values were smaller than 0.8 due to a big sampling variance (0.3) and a small population effect size (0.2). Moreover, the third level ICC strongly impacted the power. For instance, the cases with a sample size of 100 at the last block in Table 3.13 showed good power (>=0.78) with low level-3 ICC (0.05). However, the power values (<=0.41) dropped dramatically when level-3 ICC increases to 0.3. Other findings were similar to the cases with medium sample size. The third level ICC strongly influenced the power when the sample size was small, or the within-study variance was big. The results suggested that we should consider the power level when the population effect size is small in the three-level meta-regression model.

To visualize the results from the present simulation study, a heat map is displayed as Figure 3.2. There are 20 blocks in the heat map. The y axis indicates the number of studies in each group and the x axis indicates the number of groups under two population effect size. In each block, the y axis indicates the range of within-study variance from 0.05 to 0.3 (i.e., 0.05, 0.1, 0.2, 0.3), and under each within-study variance level-2 ICC shows from 0.05 to 0.3 (i.e., 0.05, 01, 0.15, 0.2, 0.25, 0.3). The x axis shows the range of level-3 ICC from 0.05 to 0.3 (i.e., 0.05, 01, 0.15, 0.2, 0.25, 0.3). The red color represents the power over 0.8, the white color represents the power around 0.8, and the blue color represents the power below 0.2.

Figure 3.2 A heat map of power values from simulation



The heat map illustrated the same results that when the sample size and population effect size went larger, more power values with different variance parameters could be considered as good power size for the weighted average effect size in the three-level meta-analysis. Higher level-3 ICC (larger between-group variance) caused lower power in the study.

Discussion

The present study in Chapter 3 extended the prior work by Hedges and Pigott (2001, 2004) for the power in meta-analysis regression model. The study followed the procedures of calculating power in two-level (random effects) mate-regression model to extend the calculation for power in three-level meta-regression model. The between-group variance was introduced to the model which indicated that possible correlations existed among studies conducted by same research group and lab. The study provided general procedures to find the variance-covariance structure for a three-level meta-regression, showed the equations to calculate non-centrality parameter in the alternative distribution, and how to use the parameter to detect statistical power in three-level meta-regression model.

The study first explored the power of the weighted average effect size in three-level meta-regression model and constructed the variance-covariance matrix used to conduct weights in the estimations. A two-level meta-regression model assumed an effect size nested within a study. The variance-covariance matrix was diagonal. The diagonal elements were a sum of two variance components - the effect size variance (which was known and varies across effect sizes) and the between-study variance, which was constant across studies and was estimated. In a three-level meta-regression model (e.g., where studies were nested within research groups), the variance-covariance matrix would be a block diagonal. The diagonal elements in each block matrix were a sum of three variance components - the effect size variance (which was known and varies across effect sizes), the between-study within-group variance which was constant across studies and was estimated, and the between-group variance with was constant across groups and was estimated. The off-diagonal elements of each block matrix were covariances between the studies linked to a specific research group. There would be as many block matrices as there are groups. The dimensions of these matrices were determined by the number of studies in each group. The study then calculated the non-centrality parameter in the alternative distribution to detect statistical power for the weighted average effect size in a three-level meta-regression model. Each model was followed by a simple illustrated example to show how to compute power statistics.

The groups were assumed as random in the third level in the present study. Potential structures could be further discussed. For instance, one condition is to test the groups at the third level are not random but fixed, which means the three-level model would flat to a two-level model. The variance-covariance matrix would change back to a diagonal matrix instead of a block matrix in the three-level model. Similar structures were discussed by Hedge and Pigott (2004) in the second level model with moderators. The weighted average effect (weighted grand mean) could be computed by calculating a weighted average of the weighted mean effect sizes from groups. Generally, compared with a three-level model, the present condition would lead to smaller variance. As a result, the non-centrality parameter would be larger, and the power of tests would be larger.

Second, the study explored the power of the moderators (individual regression coefficients) in the moderation analysis. Two variance-covariance matrix structures were shown. One assumed no random effects of the moderators in the model, and the other assumed the random effects of second-level moderators existed at the third level. The latter had the more complex variance-covariance structure. The study then calculated the non-centrality parameter in the alternative distribution to detect statistical power for the moderators in a three-level meta-regression model. Each model was followed by a simple illustrated example to show how to compute power statistics. The power statistics of the moderators for meta-regression models in the study were low because of the small regression coefficients and large standard errors. The low power of moderators in the meta-regression models were similar to the results from Hedges and Pigott's work (2004). The insufficient power can cause futile conclusion, thus the moderators with low power should be interpreted carefully. Therefore, computing power for moderators in meta-analysis seems more important than detecting power for weighted average effect size (Hedges & Pigott, 2004). In fact, the moderators or interactions in multiple regression (or called moderated multiple regression) usually suffer from a lack of power (Aguinis, 1995; Shieh, 2009). The main problem of low power is related to the product variable does potentially not distributed normally (McClelland & Judd, 1993). Therefore, to solve the problem, a transformation of the interested variables might be necessary if the variables are heavily skewed (Shieh, 2009). Structural equation modeling (SEM) was suggested as another possible solution to enhance power because the measurement error could be involved in the model (Aguinis, 1995). Based on the previous evidence, the present study suggests selecting the potential moderators from prior theories and understanding the properties of moderators before the analysis. A good practice is also to consider proper sample size and design method before conducting moderated multiple regression or moderation analysis in meta-analysis.

The simulation study showed how different values of parameters could influence the power of the weighted average effect size in a three-level meta regression model. Balanced and univariate case was considered in the current development. The values of parameters in the simulation covered a wide range of total sample sizes, the values of error terms, the levels of heterogeneity from higher levels, and the population sample sizes. Overall, the simulation study demonstrated that high level-3 ICC could cause a small value of power, which indicated the meta-analysis might lead to an invalid conclusion of the average sample size, especially in the cases with a small sample size or a large (within-study) sampling variance. The small population effect size caused small power statistics in the three-level meta-regression model. The findings also suggests that a three-level model needs a considerable sample size to ensure a good power of the meta-analysis.

The present study has some limitations. First, the current development focused on the univariate cases of three-level meta-regression models. And the simulation study used balanced cases to illustrate the final results. Future studies could extend the development to multivariate and imbalanced cases in three-level meta-regression models to capture the changes of power. Second, all parameters were assumed in the study, and they were not from empirical studies. Thus, future studies could use real examples to show how third-level heterogeneity impacts the power in the three-level meta-regression model.

APPENDICES

Appendix A Variable Summary

Table A.1 Variables extracted from ECLS-K:2011

Variables	Description	Variable name	Kindergarten	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5
Reading achievement	IRT-based scale scores	RSCALK5	X2	X4	X6	X7	X8	X9
Math achievement	IRT-based scale scores	MSCALK5	X2	X4	X6	X7	X8	X9
Science achievement	IRT-based scale scores	SSCALK5	X2	X4	X6	X7	X8	X9
ATL	Composite continuous variable with seven elements	TCHAPP	X2	X4	X6	X7	X8	X9
EPB	Composite continuous variable with five elements	TCHEXT	X2	X4	X6	X7	X8	X9
IPB	Composite continuous variable with four elements	TCHINT	X2	X4	X6	X7	X8	X9
Teacher experience	Continuous variable (unit: years)	YRSTCH	A1	A4	A6	A7	A8	A9
School enrollment	Ordinal variable was recoded to continuous variable.	KENRLS	X2	X4	X6	X7	X8	X9
School SES	Ordinal variable was recoded to continuous variable.	FRMEAL	X2FLCH2_I X4FLCH2_I	X4FMEAL_I X4RMEAL_I	X6	X7	X8	X9
Change school	Binary indicator (reference group: non-change)	DEST	X2	X4	X6	X7	X8	X9
Age	Age at spring kindergarten	X2KAGE_R						
SES	Composite continuous variable	X12SESL						
Speak non-English at home	Binary indicator (reference group: speak English at home)	X12LANGST						
Gender	Binary indicator (reference group: male)	X_CHSEX_R						
Race	Categorical variable	X_RACHTH_R	Generate four binary variables (Black students, Hispanic students, Asian students, and Other students) in the study (reference group: White students)					

Note. ATL = Approaches to Learning; EPB = Externalizing problem behaviors; IPB = Internalizing problem behaviors.

Table A.2 Descriptive statistics in complete data analysis

Year	Reading	Math	Science	ATL	EPB	IPB	Enrollment	School SES	Teacher experience
11	70.80	52.06	34.87	3.20	1.57	1.47	510.21	45.21	14.45
	14.29	13.04	7.26	0.65	0.59	0.46	216.64	28.61	9.57
12	97.59	75.13	44.38	3.14	1.69	1.52	517.95	46.11	15.01
	16.51	14.78	9.95	0.68	0.59	0.49	211.62	27.92	9.74
13	114.92	92.85	54.03	3.14	1.67	1.56	520.34	46.90	15.41
	15.67	16.96	11.19	0.68	0.59	0.50	214.41	28.10	9.71
14	123.17	106.47	61.68	3.14	1.64	1.57	513.57	46.96	14.44
	14.43	16.74	11.40	0.69	0.59	0.52	217.66	27.88	9.37
15	131.32	115.14	68.29	3.16	1.60	1.57	515.67	46.65	14.40
	13.65	16.30	11.30	0.68	0.57	0.53	223.94	28.03	9.28
16	138.44	122.20	75.10	3.18	1.59	1.55	527.32	47.10	14.29
	14.14	16.19	11.94	0.69	0.57	0.50	227.83	28.07	8.98
Total	112.71	93.98	56.39	3.16	1.63	1.54	517.51	46.49	14.67
	27.15	28.86	17.37	0.68	0.58	0.50	218.81	28.11	9.45

Year	Change school	Age	SES	Non-English	Gender	Black	Hispanic	Asian	Other
11	0.00	73.81	0.05	0.14	0.49	0.08	0.22	0.05	0.06
	0.00	4.35	0.80	0.35	0.50	0.27	0.41	0.22	0.23
12	0.04	73.81	0.05	0.14	0.49	0.08	0.22	0.05	0.06
	0.20	4.35	0.80	0.35	0.50	0.27	0.41	0.22	0.23
13	0.01	73.81	0.05	0.14	0.49	0.08	0.22	0.05	0.06
	0.10	4.35	0.80	0.35	0.50	0.27	0.41	0.22	0.23
14	0.05	73.81	0.05	0.14	0.49	0.08	0.22	0.05	0.06
	0.23	4.35	0.80	0.35	0.50	0.27	0.41	0.22	0.23

Table A.2 Continued.

15	0.05	73.81	0.05	0.14	0.49	0.08	0.22	0.05	0.06
	0.22	4.35	0.80	0.35	0.50	0.27	0.41	0.22	0.23
16	0.09	73.81	0.05	0.14	0.49	0.08	0.22	0.05	0.06
	0.29	4.35	0.80	0.35	0.50	0.27	0.41	0.22	0.23
Total	0.04	73.81	0.05	0.14	0.49	0.08	0.22	0.05	0.06
	0.20	4.34	0.80	0.35	0.50	0.27	0.41	0.22	0.23

Note. In each year, the first row indicates means of the variables and the second row indicates standard deviations of the variables. ATL = Approaches to Learning; EPB = Externalizing problem behaviors; IPB = Internalizing problem behaviors.

Table A.3 Correlation table of continuous variables (time average) in complete data analysis

	Reading	Math	Science	ATL	EPB	IPB	Enrollment	School SES	Teacher experience	Age	SES
Reading	1.00										
Math	0.69	1.00									
Science	0.65	0.68	1.00								
ATL	0.40	0.37	0.28	1.00							
EPB	-0.17	-0.14	-0.10	-0.61	1.00						
IPB	-0.17	-0.19	-0.12	-0.38	0.29	1.00					
Enrollment	0.00	-0.01	-0.03	0.02	-0.04	-0.02	1.00				
School SES	-0.28	-0.29	-0.31	-0.09	0.08	0.04	0.10	1.00			
Teacher experience	0.05	0.05	0.06	0.05	-0.05	-0.01	-0.07	-0.11	1.00		
Age	0.02	0.06	0.08	0.02	0.02	0.02	-0.04	-0.01	0.01	1.00	
SES	0.40	0.40	0.41	0.19	-0.11	-0.09	-0.04	-0.53	0.08	-0.01	1.00

Note. ATL = Approaches to Learning; EPB = Externalizing problem behaviors; IPB = Internalizing problem behaviors.

Appendix B Study Summary and Forest Plots

Table B.1 Study summary

Study	Year	Publication type	School level	Single timepoint	Longitudinal	Student type	Performance outcome	Country	Term	Included in meta-analysis
Beisly et al.	2020	Journal	Pre-k	Yes	No	Regular	Reading, Mathematics	USA	LRB	Yes
Bodovski	2007	Dissertation	1	Yes	No	Regular	Reading	USA	ATL	Yes
Brock et al.	2009	Journal	K	Yes	No	Regular	Reading, Mathematics	USA	LRB	Yes
Bumgarner et al.	2013	Journal	K, 1, 3	No	Yes	Hispanic	Mathematics	USA	ATL	No (no correlation)
Bustamante & Hindman	2019	Journal	Pre-k	Yes	Yes	Head start	Reading, Mathematics	USA	ATL	Yes
Coté	2018	Dissertation	3	No	Yes	Regular	Reading	USA	LRB	Yes
Durbrow et al.	2001	Journal	Elem	Yes	No	Remote village	Combined	the West Indies	LRB	No (combined achievement)
Durbrow et al.	2000	Journal	Elem	Yes	No	Remote village	Combined	the West Indies	LRB	No (combined achievement)
Elliott	2019	Journal	K, 3	Yes	No	Regular	Reading, Mathematics	USA	ATL	Yes
George & Greefield	2005	Journal	K, 1	Yes	Yes	Most former Head start	Combined	USA	ATL	No (combined achievement)
Jackson	2019	Dissertation	1	No	Yes	Regular	Reading, Mathematics	USA	ATL	Yes
Le et al.	2019	Journal	K	Yes	No	Regular	Advanced Mathematics	USA	ATL	No (no correlation)
Li-Grining et al.	2010	Journal	K to 6	No	Yes	Regular	Reading, Mathematics	USA	ATL	No (no correlation)
Mattews et al.	2010	Journal	K, 1, 3, 5	Yes	Yes	Regular	Reading	USA	LRB	Yes
McClelland et al.	2006	Journal	K to 6	Yes	No	Regular	Reading, Mathematics	USA	LRB	Yes (no correlation for long-term)

Table B.1 Continued

McGinnis	2009	Dissertation	3	Yes	No	Regular	Reading, Mathematics	USA	ATL	Yes
McWayne et al.	2004	Journal	Pre-k	Yes	No	Head start	Combined	USA	ATL	No (combined outcome; no correlation)
Musu-Gillette et al.	2015	Journal	k	Yes	No	Regular	Reading	USA	ATL	Yes
Neuenschwander et al.	2012	Journal	K	Yes	No	Regular	Reading, Mathematics	Switzerland	LRB	Yes
Ortiz	2014	Dissertation	Pre-k	Yes	No	Regular	Reading, Mathematics	USA	ATL	Yes
Razza et al.	2015	Journal	K, 4	Yes	Yes	Low income, Minority	Reading, Mathematics	USA	ATL	Yes
Ready et al.	2005	Journal	K	Yes	No	Regular	Reading	USA	ATL	Yes
Robinson & Mueller	2014	Journal	K	Yes	No	Regular	Mathematics	USA	ATL	Yes
Sánchez-Pérez et al.	2018	Journal	Elem	Yes	No	Regular	Reading, Mathematics	Spain	LRB	Yes
Sasser et al.	2015	Journal	Pre-k, 3	No	Yes	Head start	Reading, Mathematics	USA	LRB	Yes
Smith-Adcock et al.	2019	Journal	K	Yes	No	Low SES	Reading	USA	ATL	Yes
Stipek et al.	2010	Journal	K, 1, 3, 5	Yes	Yes	Low income	Reading	USA	LRB	Yes
Sung and Wickrama	2018	Journal	K, 1, 2	No	Yes	Regular	Reading, Mathematics	USA	ATL	Yes
Tach & Farkas	2006	Journal	K, 1	No	Yes	Regular	Reading	USA	ATL	No (no correlation)
Williams et al.	2016	Journal	Elem	No	Yes	Regular	Mathematics	Australia	ARC	Yes

Note. Elem = elementary school; LRB = Learning-related behaviors; ATL = Approaches to learning; ACR = attentional-cognitive regulation

Figure B.1 A forest plot for the relationship between ATL and reading achievement from single timepoint designs

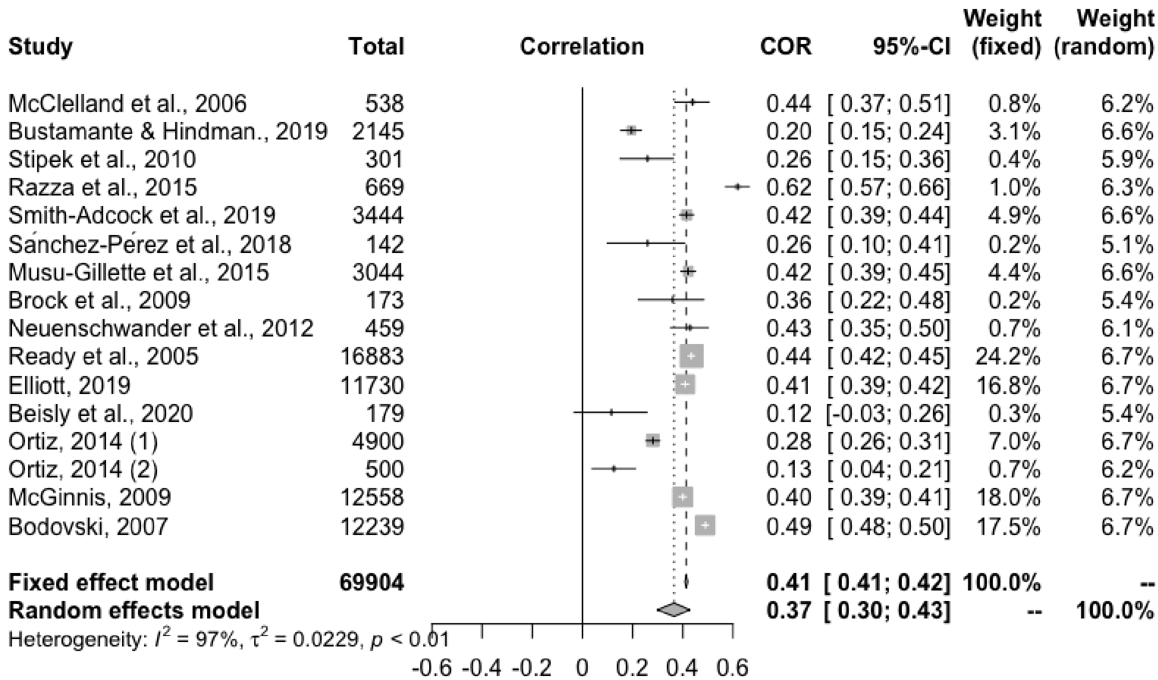


Figure B.2 A forest plot for the relationship between ATL and reading achievement from longitudinal designs

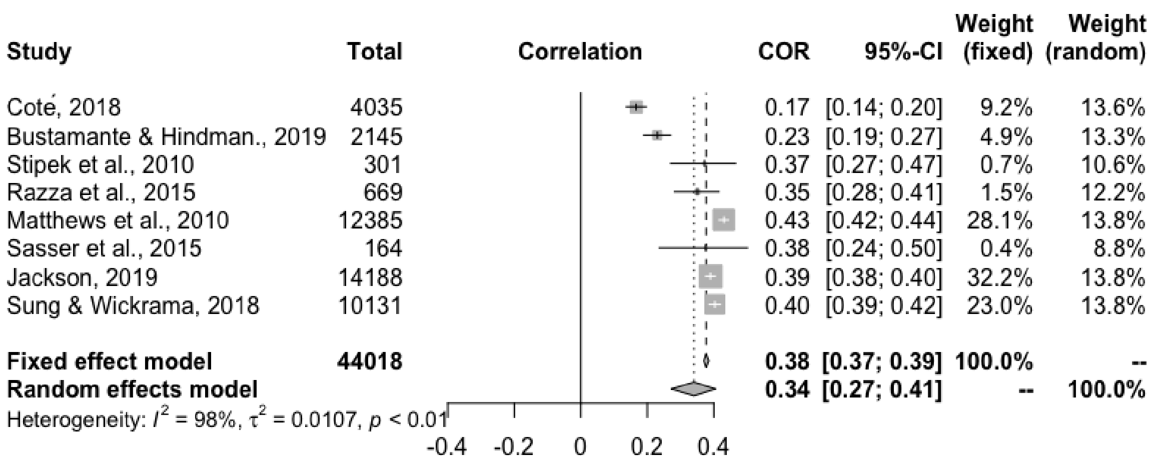


Figure B.3 A forest plot for the relationship between ATL and mathematics achievement from single timepoint designs

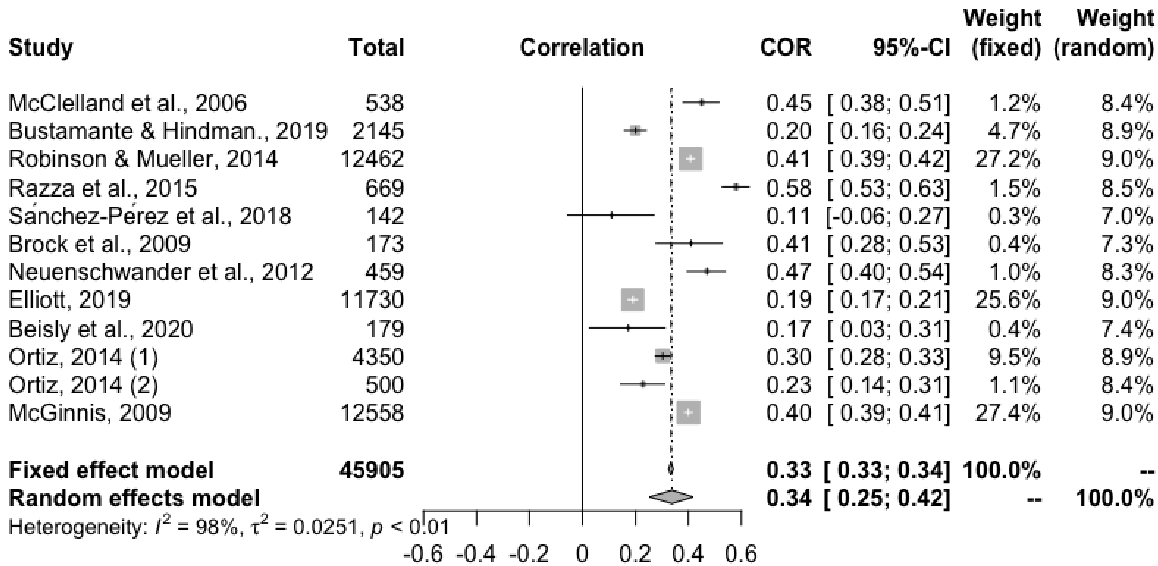
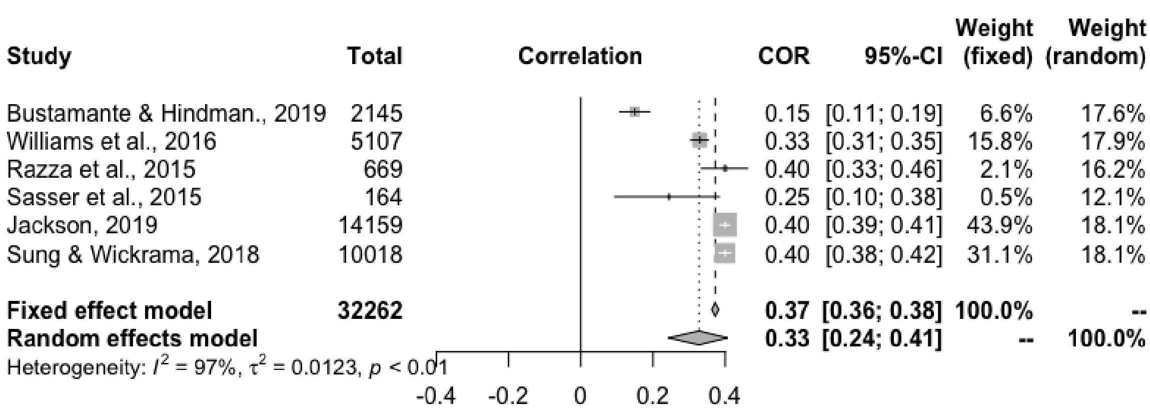


Figure B.4 A forest plot for the relationship between ATL and mathematics achievement from longitudinal designs



Appendix C Example Code

The example code in R is for the simulation study in Chapter 3. The example illustrates the results of one block with population effect size equals to 0.2, within-study variance equals to 0.3, and 6 groups in the model.

```
start.time <- Sys.time() #record start time
set.seed(12345) #set random seed

population <- 0.2 #for example population effect size is 0.2
vee <- 0.3 #for example within-study variance is 0.3
n.group <- 6 #for example 6 groups
n.study <- c(2,4,6,8,10) #number of studies per group
icc2 <- c(0.005, 0.1, 0.15, 0.2, 0.25, 0.3) #level-2 ICC
icc3 <- c(0.005, 0.1, 0.15, 0.2, 0.25, 0.3) #level-3 ICC

out <- vector("list")
for (p in 1:1000){ #1000 times iteration
  y <- vector()
  for (k in n.study){
    for (i in icc2) {
      for (j in icc3) {
        tua2 <- i/(1-i-j)*vee #between-study variance
        tua3 <- j/(1-i-j)*vee #between-group variance
        Ai = Diagonal(n=k, x=vee+tua2) + tua3 #var-cov matrix per group
        a <- list(Ai)
        group <- n.group - 1
        for(m in 1:group) {a <- c(a, Ai)}
        V3 = bdiag(a) #var-cov matrix for a three-level model
        sd12 = sqrt(tua2+vee)
        sd3 = sqrt(tua3)
        tmp2 <- rnorm(k*n.group, mean = 0, sd = sd12)
        err12 <- tmp2
        tmp3 <- rnorm(n.group, mean = 0, sd = sd3)
        err3 <- rep(tmp3, each = k)
        T <- population + err12 +err3 #observed effect sizes
        Z <- rep(1, k*n.group)
        W <- solve(V3) #weight matrix
        A <- t(Z)%*%W%*%Z
        B <- t(Z)%*%W%*%T
        lambda <-(1/A)*B/sqrt(1/A) #non-centriality parameter of z test
        lambda <- as.numeric(lambda)
```

```
        #power of the weighted average effect size, two-tailed test, type I error = 0.05
        power <- 1-pnorm(1.96-lambda)+pnorm(-1.96-lambda)
        y<-c(y, power)
      }
    }
  }
  out[[p]] <- as.matrix(y)
}
df <- data.frame(matrix(unlist(out), ncol = max(lengths(out)), byrow = TRUE))
average <- colMeans(df) #take average
ID <- rep(1:6, 30) #reframe simulated results
dataframe<- data.frame(ID,average)
dataframe<-unstack(dataframe, average~ID)

end.time <- Sys.time() #record ending time
time.taken <- end.time - start.time
time.taken #compute running time
```

REFERENCES

REFERENCES

References marked with an asterisk indicate studies included in Table B.1.

Aguinis, H. (1995). Statistical power problems with moderated multiple regression in management research. *Journal of Management*, *21*(6), 1141-1158.

Alexander, R. A. (1990). A note on averaging correlations. *Bulletin of the Psychonomic Society*, *28*(4), 335-336.

Algozzine, B., Wang, C., & Violette, A. S. (2011). Reexamining the relationship between academic achievement and social behavior. *Journal of Positive Behavior Interventions*, *13*(1), 3-16.

Anderson, S., & Messick, S. (1974). Social competency in young children. *Developmental Psychology*, *10*(2), 282-293.

Anthony, C. J., DiPerna, J. C., & Amato, P. R. (2014). Divorce, approaches to learning, and children's academic achievement: A longitudinal analysis of mediated and moderated effects. *Journal of School Psychology*, *52*(3), 249-261.

Anthony, C. J., & Ogg, J. (2019). Parent involvement, approaches to learning, and student achievement: Examining longitudinal mediation. *School Psychology*, *34*(4), 376-385.

Atkins-Burnett, S. (2007). *Measuring children's progress from preschool through third grade*. Mathematica Policy Research.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173-1182.

Becker, B. J. (2000). Multivariate meta-analysis. *Handbook of applied multivariate statistics and mathematical modeling*, 499-525.

*Beisly, A., Kwon, K. A., Jeon, S., & Lim, C. (2020). The moderating role of two learning related behaviours in preschool children's academic outcomes: learning behaviour and executive function. *Early Child Development and Care*, 1-16.

Birch, S., & Ladd, G. (1997). The teacher–child relationship and children's early school adjustment. *Journal of School Psychology, 35*, 61-79.

*Bodovski, K. (2010). Concerted cultivation, students' behavior, and unequal achievement in elementary school (Order No. AAI3380714) [Doctoral dissertation, The Pennsylvania State University]. APA PsycInfo.

Bodovski, K., & Farkas, G. (2007). Mathematics growth in early elementary school: The roles of beginning knowledge, student engagement, and instruction. *The Elementary School Journal*, *108*(2), 115-130.

Bodrova, E., & Leong, D. J. (1996). Tools of the Mind: The Vygotskian Approach to Early Childhood Education. *Englewood Cliffs, New Jersey: Merrill, an imprint of Prentice Hall*.

Bodrova, E., & Leong, D. J. (2019). Tools of the Mind: The Vygotskian-Based Early Childhood Program. *Journal of Cognitive Education and Psychology*, *17*(3), 223-237.

Boedeker, P. (2017). Hierarchical linear modeling with maximum likelihood, restricted maximum likelihood, and fully Bayesian estimation. *Practical Assessment, Research, and Evaluation*, *22*(1), 2.

Borenstein, M., Hedges, L., & Rothstein, H. (2007). Meta-analysis: Fixed effect vs. random effects. *Meta-analysis. com.*

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2021). *Introduction to meta-analysis*. John Wiley & Sons.

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

Bulotsky-Shearer, R. J., Fernandez, V., Dominguez, X., & Rouse, H. L. (2011). Behavior problems in learning activities and social interactions in Head Start classrooms and early reading, mathematics, and approaches to learning. *School Psychology Review*, *40*(1), 39-56.

*Bumgarner, E., Martin, A., & Brooks-Gunn, J. (2013). Approaches to learning and Hispanic children's math scores: The moderating role of English proficiency. *Hispanic Journal of Behavioral Sciences*, *35*(2), 241-259.

*Bustamante, A. S., & Hindman, A. H. (2019). Classroom quality and academic school readiness outcomes in Head Start: The indirect effect of approaches to learning. *Early Education and Development*, *30*(1), 19-35.

Bustamante, A. S., White, L. J., & Greenfield, D. B. (2017). Approaches to learning and school readiness in head start: Applications to preschool science. *Learning and Individual Differences, 56,* 112–118.

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365-376.

Campbell, S. B. (1995). Behavior problems in preschool children: A review of recent research. *Journal of Child Psychology and Psychiatry*, *36*(1), 113-149.

Cafri, G., Kromrey, J. D., & Brannick, M. T. (2010). A meta-meta-analysis: Empirical review of statistical power, type I error rates, effect sizes, and model selection of meta-analyses published in psychology. *Multivariate Behavioral Research*, *45*(2), 239-270.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (2nd Ed.). New York: Academic Press.

Comulada, W. S. (2015). Model specification and bootstrapping for multiply imputed data: An application to count models for the frequency of alcohol use. *The Stata Journal*, *15*(3), 833-844.

Cooper, D. H., & Farran, D. C. (1991). *The Cooper-Farran behavioral rating scales*. Clinical Psychology Publishing Company, Incorporated.

Corbeil, R. R., & Searle, S. R. (1976). A comparison of variance component estimators. *Biometrics*, 779-791.

Costa, A. L., & Kallick, B. (Eds.). (2008). *Learning and leading with habits of mind: 16 essential characteristics for success*. ASCD.

*Coté, P. S. (2019). Examining the predictive capacity of a set of learning related social skills inkindergarten on the academic achievement in reading in third grade (Order No. AAI10978148) [Doctoral dissertation, Tarleton State University]. APA PsycInfo.

Diamond, A., Barnett, W. S., Thomas, J., & Munro, S. (2007). Preschool program improves cognitive control. *Science (New York, NY), 318*(5855), 1387.

Dignath, C., Buettner, G., & Langfeldt, H. P. (2008). How can primary school students learn self-regulated learning strategies most effectively?: A meta-analysis on self-regulation training programmes. *Educational Research Review*, *3*(2), 101-129.

DiPerna, J. C. (2006). Academic enablers and student achievement: Implications for assessment and intervention services in the schools. *Psychology in the Schools, 43*(1), 7-17.

DiPerna, J. C., & Elliott, S. N. (2002). Promoting academic enablers to improve student achievement: An introduction to the mini-series. *School Psychology Review*, *31*(3), 293-297.

DiPerna, J. C., Volpe, R. J., & Elliott, S. N. (2002). A model of academic enablers and elementary reading/language arts achievement. *School Psychology Review, 31*, 298- 312.

Duff, A., Boyle, E., Dunleavy, K., & Ferguson, J. (2004). The relationship between personality, approach to learning and academic performance. *Personality and Individual Differences, 36*(8), 1907-1920.

*Durbrow, E. H., Schaefer, B. A., & Jimerson, S. R. (2000). Learning behaviours, attention and anxiety in Caribbean children: Beyond the 'usual suspects' in explaining academic performance. *School Psychology International*, *21*(3), 242-251.

*Durbrow, E. H., Schaefer, B. A., & Jimerson, S. R. (2001). Learning-related behaviours versus cognitive ability in the academic performance of Vincentian children. *British Journal of Educational Psychology*, *71*(3), 471-483.

Efron, B., & Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, 54-75.

Fernández-Castilla, B., Jamshidi, L., Declercq, L., Beretvas, S. N., Onghena, P., & Van den Noortgate, W. (2020). The application of meta-analytic (multi-level) models with multiple random effects: A systematic review. *Behavior Research Methods*, *52*, 2031-2052.

Fisher, R. A. (1921). On the 'probable error' of a coefficient of correlation deduced from a small sample. *Metron*, *1*, 1-32.

*George, J. L., & Greenfield, D. B. (2005). Examination of a structured problem-solving flexibility task for assessing approaches to learning in young children: Relation to teacher ratings and children's achievement. *Journal of Applied Developmental Psychology*, *26*(1), 69-84.

Giesselmann, M., & Schmidt-Catran, A. W. (2020). Interactions in fixed effects regression models. *Sociological Methods & Research*, 1-28.

Gresham, F. M., & Elliott, S. N. (1990). *Social skills rating system: Manual*. American Guidance Service.

Hattie, J., Biggs, J., & Purdie, N. (1996). Effects of learning skills interventions on student learning: A meta-analysis. *Review of Educational Research*, *66*(2), 99-136.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica: Journal of the Econometric Society*, 1251-1271.

Hedges, L. V. (1992). Meta-analysis. *Journal of Educational Statistics*, *17*(4), 279-296.

Hedges, L., & Olkin, I. (2014). *Statistical methods for meta-analysis*. Academic press.

Hedges, L. V., & Pigott, T. D. (2001). The power of statistical tests in meta-analysis. *Psychological Methods*, *6*(3), 203-217.

Hedges, L. V., & Pigott, T. D. (2004). The power of statistical tests for moderators in meta-analysis. *Psychological Methods, 9*(4), 426-445.

Hedges, L. V., & Vevea, J. L. (1998). Fixed-and random-effects models in meta-analysis. *Psychological Methods, 3*(4), 486-504.

Higgins, J. P., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*(11), 1539-1558.

Hong, G., & Raudenbush, S. W. (2005). Effects of kindergarten retention policy on children's cognitive growth in reading and mathematics. *Educational Evaluation and Policy Analysis*, *27*(3), 205-224.

Hong, G., & Yu, B. (2007). Early-grade retention and children's reading and math learning in elementary years. *Educational Evaluation and Policy Analysis*, *29*(4), 239-261.

Howse, R. B., Calkins, S. D., Anastopoulos, A. D., Keane, S. P., & Shelton, T. L. (2003). Regulatory contributors to children's kindergarten achievement. *Early Education and Development*, *14*(1), 101-120.

*Jackson, G. N. (2019). The impact of kindergarten students' executive functions on academic success: A secondary data analysis on at-risk children (Order No. AAI13897396) [Doctoral dissertation, Fielding Graduate University]. APA PsycInfo.

Jackson, D., Riley, R., & White, I. R. (2011). Multivariate meta-analysis: potential and promise. *Statistics in Medicine, 30*(20), 2481-2498.

Kagan, S. L., Moore, E. K., & Bredekamp, S. (1995). *Reconsidering children's early development and learning: Toward common views and vocabulary* (Vol. 95, No. 3). National Education Goals Panel.

Katz, L.G. (1985). Dispositions in early childhood education. *ERIC/EECE Bulletin, 18*(2), 1-3.

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

Konstantopoulos, S. (2008). The power of the test for treatment effects in three-level cluster randomized designs. *Journal of Research on Educational Effectiveness, 1*(1), 66-88.

Langan, D., Higgins, J. P., Jackson, D., Bowden, J., Veroniki, A. A., Kontopantelis, E., ... & Simmonds, M. (2019). A comparison of heterogeneity variance estimators in simulated random-effects meta-analyses. *Research Synthesis Methods*, *10*(1), 83-98.

*Le, V. N., Schaack, D., Neishi, K., Hernandez, M. W., & Blank, R. (2019). Advanced content coverage at kindergarten: Are there trade-offs between academic achievement and social-emotional skills?. *American Educational Research Journal*, *56*(4), 1254-1280.

*Li-Grining, C. P., Votruba-Drzal, E., Maldonado-Carreño, C., & Haas, K. (2010). Children's early approaches to learning and academic trajectories through fifth grade. *Developmental Psychology*, *46*(5), 1062.

Longford, N. T. (1987). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. *Biometrika*, *74*(4), 817-827.

Malecki, C. K., & Elliot, S. N. (2002). Children's social behaviors as predictors of academic achievement: A longitudinal analysis. *School Psychology Quarterly*, *17*(1), 1-23.

*Matthews, J. S., Kizzie, K. T., Rowley, S. J., & Cortina, K. (2010). African Americans and boys: Understanding the literacy gap, tracing academic trajectories, and evaluating the role of learning-related skills. *Journal of Educational Psychology*, *102*(3), 757-771.

McClelland, G. H., & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin*, *114*(2), 376-390.

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

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

McCormick, M. P., Cappella, E., O'Connor, E., Hill, J. L., & McClowry, S. (2016). Do effects of social-emotional learning programs vary by level of parent participation? Evidence from the randomized trial of INSIGHTS. *Journal of Research on Educational Effectiveness*, *9*(3), 364-394.

McDermott, P. A. (1999). National scales of differential learning behaviors among American children and adolescents. *School Psychology Review*, *28*(2), 280-291.

McDermott, P. A., Green, L. F., Francis, J. M., & Stott, D. H. (2000). Preschool Learning Behavior Scale. *School Psychology Quarterly*.

McDermott, P. A., Leigh, N. M., & Perry, M. A. (2002). Development and validation of the preschool learning behaviors scale. *Psychology in the Schools*, *39*(4), 353-365.

McDermott, P. A., Rikoon, S. H., Waterman, C., & Fantuzzo, J. W. (2012). The preschool learning behaviors scale: dimensionality and external validity in head start. *School Psychology Review*, *41*(1), 66-81.

*McGinnis, A. M. (2010). Student behaviors as predictors of later academic achievement: School entry through fifth grade (Order No. AAI3380966) [Doctoral dissertation, The Pennsylvania State University]. APA PsycInfo.

*McWayne, C. M., Fantuzzo, J. W., & McDermott, P. A. (2004). Preschool competency in context: an investigation of the unique contribution of child competencies to early academic success. *Developmental Psychology*, *40*(4), 633.

Mega, C., Ronconi, L., & De Beni, R. (2014). What makes a good student? How emotions, self-regulated learning, and motivation contribute to academic achievement. *Journal of Educational Psychology*, *106*(1), 121-131.

Meisels, S. J., Atkins-Burnett, S., & Nicholson, J. (1996). *Assessment of social competence, adaptive behaviors, and approaches to learning with young children*. US Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

Meng, C. (2015). Classroom quality and academic skills: Approaches to learning as a moderator. *School Psychology Quarterly*, *30*(4), 553-563.

Miller, N., & Pollock, V. E. (1994). Meta-analytic synthesis for theory development. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (p. 457–483). Russell Sage Foundation.

Mokrova, I. L., O'Brien, M., Calkins, S. D., Leerkes, E. M., & Marcovitch, S. (2013). The role of persistence at preschool age in academic skills at kindergarten. *European Journal of Psychology of Education*, *28*(4), 1495-1503.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica: Kournal of the Econometric Society*, 69-85.

*Musu-Gillette, L. E., Barofsky, M. Y., & List, A. (2015). Exploring the relationship between student approaches to learning and reading achievement at the school level. *Journal of Early Childhood Literacy*, *15*(1), 37-72.

Najarian, M., Tourangeau, K., Nord, C., and Wallner-Allen, K. (2018a). *Early Childhood Longitudinal Study, Kindergarten Class of 2010–11 (ECLS-K:2011), Kindergarten Psychometric Report* (NCES 2018-182). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Retrieved from http://nces.ed.gov/pubsearch.

Najarian, M., Tourangeau, K., Nord, C., and Wallner-Allen, K. (2018b). *Early Childhood Longitudinal Study, Kindergarten Class of 2010–11 (ECLS-K:2011), First- and Second-Grade Psychometric Report (NCES 2018-183).* National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Retrieved from https://nces.ed.gov/pubsearch.

Najarian, M., Tourangeau, K., Nord, C., Wallner-Allen, K., and Vaden-Kiernan, N. (2019). *Early Childhood Longitudinal Study, Kindergarten Class of 2010–11 (ECLS-K:2011), Third-Grade, Fourth-Grade, and Fifth-Grade Psychometric Report* (NCES 2020-123). U.S. Department of Education. Institute of Education Sciences. Washington, DC: National Center for Education Statistics. Retrieved from https://nces.ed.gov/pubsearch.

National Center for Education Statistics. (n.d.). Comparisons to the ECLS-B & ECLS-K. Retrieved from: https://nces.ed.gov/ecls/comparisons2011.asp

National Education Goals Panel. (1991). The national education goals report. Washington, DC. U.S. Department of Education.

Nelson, J. R., Benner, G. J., Lane, K., & Smith, B. W. (2004). Academic achievement of K-12 students with emotional and behavioral disorders. *Exceptional Children*, *71*(1), 59-73.

Nesbitt, K. T., Farran, D. C., & Fuhs, M. W. (2015). Executive function skills and academic achievement gains in prekindergarten: Contributions of learning-related behaviors. *Developmental Psychology*, *51*(7), 865-878.

*Neuenschwander, R., Röthlisberger, M., Cimeli, P., & Roebers, C. M. (2012). How do different aspects of self-regulation predict successful adaptation to school?. *Journal of Experimental Child Psychology*, *113*(3), 353-371.

Olkin, I., & Pratt, J. W. (1958). Unbiased estimation of certain correlation coefficients. *The Annals of Mathematical Statistics*, *29*(1), 201-211.

*Ortiz, A. (2016). Effects of the home environment, social skills, and approaches to learning on school readiness in children based on primary language spoken at home (Order No. AAI3690148) [Doctoral dissertation, The Pennsylvania State University]. APA PsycInfo.

Peugh, J. L. (2010). A practical guide to multilevel modeling. *Journal of School Psychology*, *48*(1), 85-112.

Rabiner, D. L., Malone, P. S., & Conduct Problems Prevention Research Group. (2004). The impact of tutoring on early reading achievement for children with and without attention problems. *Journal of Abnormal Child Psychology*, *32*(3), 273-284.

Ramdass, D., & Zimmerman, B. J. (2011). Developing self-regulation skills: The important role of homework. *Journal of Advanced Academics*, *22*(2), 194-218.

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

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congfon Jr, R., T., & Du Toit, M. (2019). *HLM 8: Hierarchical linear and nonlinear modeling*. Scientific Software International.

*Razza, R. A., Martin, A., & Brooks-Gunn, J. (2015). Are approaches to learning in kindergarten associated with academic and social competence similarly?. *Child & Youth Care Forum*, *44*, 757-776.

Ready, D. D., LoGerfo, L. F., Burkam, D. T., & Lee, V. E. (2005). Explaining girls' advantage in kindergarten literacy learning: Do classroom behaviors make a difference?. *The Elementary School Journal*, *106*(1), 21-38.

Rhoades, B. L., Warren, H. K., Domitrovich, C. E., & Greenberg, M. T. (2011). Examining the link between preschool social–emotional competence and first grade academic achievement: The role of attention skills. *Early Childhood Research Quarterly*, *26*(2), 182-191.

Rikoon, S. H., McDermott, P. A., & Fantuzzo, J. W. (2012). Approaches to learning among Head Start alumni: Structure and validity of the Learning Behaviors Scale. *School Psychology Review*, *41*(3), 272-294.

Robinson, K. (2013). Early disparities in mathematics gains among poor and non-poor children: Examining the role of behavioral engagement in learning. *The Elementary School Journal*, *114*(1), 22-47.

*Robinson, K., & Mueller, A. S. (2014). Behavioral engagement in learning and math achievement over kindergarten: A contextual analysis. *American Journal of Education*, *120*(3), 325-349.

Rosenthal, R. (1994). Parametric measures of effect size. In H.M. Cooper & L.V. Hedges (Eds.), The handbook of research synthesis (pp. 231-244). New York: Sage.

Rosenthal, R. (1995). Writing meta-analytic reviews. *Psychological Bulletin, 118*, 183-192.

*Sánchez-Pérez, N., Fuentes, L. J., Eisenberg, N., & González-Salinas, C. (2018). Effortful control is associated with children's school functioning via learning-related behaviors. *Learning and Individual Differences*, *63*, 78-88.

*Sasser, T. R., Bierman, K. L., & Heinrichs, B. (2015). Executive functioning and school adjustment: The mediational role of pre-kindergarten learning-related behaviors. *Early Childhood Research Quarterly*, *30*, 70-79.

Schaefer, B. A., Shur, K. F., Macri-Summers, M., & MacDonald, S. L. (2004). Preschool children's learning behaviors, concept attainment, social skills, and problem behaviors: validity evidence for preschool learning behaviors scale scores. *Journal of Psychoeducational Assessment*, *22*(1), 15-32.

Schafer, J. L. (1999). Multiple imputation: a primer. *Statistical Methods in Medical Research*, *8*(1), 3-15.

Schmitt, S. A., McClelland, M. M., Tominey, S. L., & Acock, A. C. (2015). Strengthening school readiness for Head Start children: Evaluation of a self-regulation intervention. *Early Childhood Research Quarterly*, *30*, 20-31.

Schomaker, M., & Heumann, C. (2018). Bootstrap inference when using multiple imputation. *Statistics in Medicine*, *37*(14), 2252-2266.

Schunck, R. (2013). Within and between estimates in random-effects models: Advantages and drawbacks of correlated random effects and hybrid models. *The Stata Journal*, *13*(1), 65-76.

Shieh, G. (2009). Detecting interaction effects in moderated multiple regression with continuous variables power and sample size considerations. *Organizational Research Methods*, *12*(3), 510-528.

Silver, N. C., & Dunlap, W. P. (1987). Averaging correlation coefficients: should Fisher's z transformation be used?. *Journal of Applied Psychology*, *72*(1), 146-148.

*Smith-Adcock, S., Leite, W., Kaya, Y., & Amatea, E. (2019). A model of parenting risk and resilience, social-emotional readiness, and reading achievement in kindergarten children from low-income families model. *Journal of Child and Family Studies*, *28*(10), 2826-2841.

*Stipek, D., Newton, S., & Chudgar, A. (2010). Learning-related behaviors and literacy achievement in elementary school-aged children. *Early Childhood Research Quarterly*, *25*(3), 385-395.

Strube, M. J. (1988). Averaging correlation coefficients: Influence of heterogeneity and set size. *Journal of Applied Psychology*, *73*(3), 559-568.

*Sung, J., & Wickrama, K. A. (2018). Longitudinal relationship between early academic achievement and executive function: Mediating role of approaches to learning. *Contemporary Educational Psychology*, *54*, 171-183.

*Tach, L. M., & Farkas, G. (2006). Learning-related behaviors, cognitive skills, and ability grouping when schooling begins. *Social Science Research*, *35*(4), 1048-1079.

Tourangeau, K., Nord, C., Lê, T., Wallner-Allen, K., Vaden-Kiernan, N., Blaker, L., & Najarian, M. (2018). Early Childhood Longitudinal Study, Kindergarten Class of 2010-11 (ECLS-K: 2011): User's Manual for the ECLS-K: 2011 Kindergarten-Fourth Grade Data File and Electronic Codebook, Public Version. NCES 2018-032. *National Center for Education Statistics*.

U.S. Department of Health and Human Services. (2019). Head Start early learning outcomes framework. Retrieved from: https://montessoriadvocacy.org/wp-content/uploads/2019/10/ Head-Start-Early-Learning-Framework.pdf

Valentine, J. C., Pigott, T. D., & Rothstein, H. R. (2010). How many studies do you need? A primer on statistical power for meta-analysis. *Journal of Educational and Behavioral Statistics*, *35*(2), 215-247.

Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2013). Three-level meta-analysis of dependent effect sizes. *Behavior Research Methods*, *45*(2), 576-594.

Veroniki, A. A., Jackson, D., Viechtbauer, W., Bender, R., Bowden, J., Knapp, G., ... & Salanti, G. (2016). Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Research Synthesis Methods*, *7*(1), 55-79.

Viana, M. A. (1982). Combined estimators for the correlation coefficient. *Communications in Statistics-Theory and Methods*, *11*(13), 1483-1504.

Vitiello, V. E., Greenfield, D. B., Munis, P., & George, J. L. (2011). Cognitive flexibility, approaches to learning, and academic school readiness in Head Start preschool children. *Early Education & Development*, *22*(3), 388-410.

Wibrowski, C. R., Matthews, W. K., & Kitsantas, A. (2017). The role of a skills learning support program on first-generation college students' self-regulation, motivation, and academic achievement: A longitudinal study. *Journal of College Student Retention: Research, Theory & Practice*, *19*(3), 317-332.

*Williams, K. E., White, S. L., & MacDonald, A. (2016). Early mathematics achievement of boys and girls: Do differences in early self-regulation pathways explain later achievement?. *Learning and Individual Differences*, *51*, 199-209.

Wooldridge, J. M. (2005). Fixed-effects and related estimators for correlated random-coefficient and treatment-effect panel data models. *Review of Economics and Statistics*, *87*(2), 385-390.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

Zins, J. E. (Ed.). (2004). *Building academic success on social and emotional learning: What does the research say?*. Teachers College Press.