Graduate School ETD Form 9
(Revised 12/07)

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Saiying Steenbergen-Hu

Entitled: The effects of acceleration on high-ability learners: A meta-analysis

For the degree of Doctor of Philosophy

Is approved by the final examining committee:

Sidney Moon
Chair

Marcia Gentry

Rebecca Mann

Yan Ping Xin

To the best of my knowledge and as understood by the student in the *Research Integrity and Copyright Disclaimer (Graduate School Form 20)*, this thesis/dissertation adheres to the provisions of Purdue University's "Policy on Integrity in Research" and the use of copyrighted material.

Approved by Major Professor(s): Sidney Moon

Approved by: A. G. Rud — March 5th, 2009
Head of the Graduate Program — Date

Graduate School Form 20
(Revised 6/09)

PURDUE UNIVERSITY
GRADUATE SCHOOL

Research Integrity and Copyright Disclaimer

Title of Thesis/Dissertation:
The effects of acceleration on high-ability learners: A meta-analysis

For the degree of Doctor of Philosophy

I certify that in the preparation of this thesis, I have observed the provisions of *Purdue University Executive Memorandum No. C-22,* September 6, 1991, *Policy on Integrity in Research.**

Further, I certify that this work is free of plagiarism and all materials appearing in this thesis/dissertation have been properly quoted and attributed.

I certify that all copyrighted material incorporated into this thesis/dissertation is in compliance with the United States' copyright law and that I have received written permission from the copyright owners for my use of their work, which is beyond the scope of the law. I agree to indemnify and save harmless Purdue University from any and all claims that may be asserted or that may arise from any copyright violation.

Saiying Steenbergen-Hu
Printed Name and Signature of Candidate

July 1, 2009
Date (month/day/year)

*Located at http://www.purdue.edu/policies/pages/teach_res_outreach/c_22.html

THE EFFECTS OF ACCELERATION ON HIGH-ABILITY LEARNERS: A META-ANALYSIS

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Saiying Steenbergen-Hu

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

August 2009

Purdue University

West Lafayette, Indiana

UMI Number: 3379769

To my late mother Mingyu Wang, my father Wujia Hu, and my husband John Steenbergen, for their love and support.

ACKNOLEDGEMENTS

I would like to express my appreciation to many people who supported me and encouraged me to come to this point. First, I owe special thanks to my advisor and committee chair Dr. Sidney Moon. Dr. Moon has been a great mentor to me since August 2004. I thank Dr. Moon for nominating me for the Purdue University Ross Fellowship. This made my study at Purdue University possible in the first place. Because of this fellowship, I had great opportunities to collaborate with her and other professors on several research projects. I gained precious research experience from these collaborations. Dr. Moon is an excellent scholar with extensive expertise in educational psychology and gifted education. A rigorous attitude and a professional manner toward scholarly work were the best things I learned from her. I am always grateful that Dr. Moon was able to provide very helpful feedback to my work on time, sometimes working on weekends and during vacations. She has nurtured my professional development in great measure.

Second, I sincerely thank my committee: Dr. Marcia Gentry, Rebecca Mann and Yan Ping Xin. I thank them for their intelligent questions and comments on my dissertation. Together with my committee chair, they played an important role in helping me to improve my work and leverage my ability as a scholar. I especially appreciate the support from Dr. Marcia Gentry. Dr. Gentry generously provided me terrific opportunities to collaborate with her on several research projects during the three years when I worked as her research assistant. She also encouraged and provided tremendous help to me and other graduate students to participate national and international conferences. The experience at conferences is an important part of learning experience as a doctoral student.

Third, I would like to acknowledge some scholars outside of my committee, including Dr. Jean Peterson, Harris Cooper, William Asher, and Oliver Wendt. They provided invaluable assistance and guidance regarding research methods and meta-analysis. I also would like to thank all the authors of primary research who made my meta-analysis possible. My thanks go to Dr. Sally Reis, Dr. Rebecca Eckert, Dr. Betsy McCoach, Dr. Joan Jacobs, Dr. Michael Coyne, Dr. Micheal Sayler, Dr. Ryan Wells, Dr. David Lohman, Dr. Maureen Marron, and Dr. Karen Rogers for providing me with their most recent studies on acceleration. I thank Dr. David Yun Dai from SUNY New York, for providing me a grant work opportunity in my fifth year of study.

Fourth, I wish to acknowledge my fellow graduate students at Purdue University for their precious encouragement and help, especially Yoona, Iryna and Dake Zhang. Special thanks also go to my old friends and colleagues back in China, including Jie Chang, Hongying Wan, Fei Xu, Xia Sun, Huizheng Gu, and Yanhua Hou. They have been tremendously supportive of me during these years of study. Their invaluable support and encouragement has lifted my heart and kept me going.

Finally, my deepest thanks go to my family. I thank my late mother, who taught me the value of hard-work, persistence, and kindness. I thank my father, sisters, and brothers for understanding me in my hard times and for sharing my joy in the good times. I feel extremely fortunate to be with my husband John Steenbergen. He is very loving, caring and supportive. He is always the first reader of my writing. He is happy to go through hundreds pages of my work and help me with grammar and composition. We had great fun going through my writing together. He is indeed a terrific partner. Thanks also go to my parents-in-law, my sister-in-law Margie and her husband Stephen. They have been very helpful and supportive of my study.

TABLE OF CONTENTS

Page

Page

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Steenbergen-Hu, Saiying. Ph. D., Purdue University, August 2009. The effects of acceleration on high-ability learners: A meta-analysis. Major Professor: Sidney Moon.

Current empirical research findings about the effects of acceleration on high-ability learners' academic achievement and social-emotional development were synthesized using meta-analytic techniques. A total of 38 primary studies conducted between 1984 and 2008 were included. The included studies were closely examined to ensure that accelerated high-ability learners were compared with appropriate comparison groups. Hedges's *g* was used as the primary effect size index. Analyses were performed using random effects models, which assume that the effects vary across different contexts, intervention conditions, and /or subjects. The overall effects of acceleration were analyzed first. Then, the results were broken down by developmental levels (P-12 and postsecondary) and comparison groups (whether accelerants were compared with same age, older age, or mixed-age peers). In addition, analyses were conducted to identify potential moderators of the effects. Results were interpreted in terms of practical significance and were also compared with those from relevant previous meta-analytic studies.

In terms of academic achievement effects, the findings from this meta-analysis are consistent with the conclusions from previous meta-analytic studies, suggesting that acceleration had a positive impact on high-ability learners. When the academic achievement effects were sorted by developmental levels, positive effects were found at both levels. The sub-group of 'with same age peers'

consistently showed a positive effect on academic achievement that were higher than the other subgroups, suggesting that the effects of acceleration may be more discernable when accelerated high-ability learners are compared with their non-accelerated same age peers. Furthermore, acceleration duration and statistical analysis were identified as moderators of academic achievement effects.

The effects of acceleration on high-ability learners' social-emotional development appeared to be slightly positive, although the positive effect was not as strong as for academic achievement. However, compared to prior meta-analytic studies, a more positive impression of the effects of acceleration on social-emotional development was found.

CHAPTER 1 INTRODUCTION

This chapter introduces the background of the study and specifies research questions. It then describes the rationale for the study and addresses the advantages and disadvantages of meta-analysis. Finally, conceptual and operational definitions of key terms of this study are provided.

1.1 Statement of the Problem

Acceleration is defined as a type of educational intervention based on progress through educational programs either at rates faster than, or at ages younger than, peers (Pressey, 1949). As experts from NAGC (1992) noted, acceleration practices aim to provide the appropriate level of challenge to students and shorten the time to complete traditional schooling. In other words, acceleration is intended to provide appropriate curriculum and services to gifted learners at a level commensurate with their demonstrated readiness and need (VanTassel-Baska, 1992). According to Tannenbaum (1958), the earliest program of acceleration was implemented by St. Louis public schools in 1862. Acceleration is commonly referred to as academic acceleration (Southern & Jones, 1991).

Acceleration has been one of the most important issues in gifted education. Its effects, particularly on student achievement and social-emotional development, have been extensively studied. A growing body of research reviews exists focusing on integrating the results of acceleration. Research reviews, and especially narrative reviews, have revealed a positive influence from acceleration on the high-ability learners' academic achievement. For example, *A Nation Deceived: How Schools Hold Back America's Brightest*

Students (i.e., The Templeton National Report on Acceleration) (Colangelo, Assouline, & Gross, 2004) presents a comprehensive review of acceleration. With the support of experts on acceleration, the authors of this report concluded that acceleration is the most effective and efficient intervention option for high-ability learners and has long-term positive effects on them.

These previous narrative reviews have provided important information for further research and development of acceleration practices. However, the utilization of the findings from these reviews is questionable. First, there is concern about the validity of the narrative review method. In fact, the traditional narrative review has been constantly criticized. For instance, Kulik (2004) noted that reviewers who use traditional narrative review methods often fail to ensure comprehensive literature searches, complete treatment of study findings, or clear identification of the relationships between study findings and review conclusions, so that they leave themselves open to charges of bias and subjectivity. As a result, the conclusions from these reviews lack credibility and can be easily dismissed.

Second, the knowledge and information gained from the traditional review are always not practically-oriented to influence educational policy decision-making. As Rich (1983) argued:

> Traditional synthesis is oriented toward providing ‘state of the art’ literature reviews; in many cases they focus on ‘objectively’ summarizing what is known about an area of scholarly literature. These ‘state of the art’ reviews are inherently not oriented toward specific policy making or administrative agendas of officials responsible for day-to-day decision making….[K]knowledge synthesis must be oriented toward utilization (p.294).

Third, the effects or conclusions from traditional narrative reviews are usually fairly fragmentary. As a result, there is a danger that the aggregated knowledge would be misused in some unintended situations and lead to harmful results (Hedges & Waddington, 1993). On the other hand, education researchers have claimed that in many cases the effects of educational factors are most likely to be context-dependent. Hedges and Waddington (1993) argued that, for the purpose of influencing educational policy decisions, researchers should identify specific variables to be changed and provide the quantified effects of the change in the related contexts.

Since the advent of meta-analysis, quantitative methods of synthesizing research have grown greatly in various fields. Meta-analysis has not only been considered to be a supplement to traditional narrative review, but also as a comparatively new and advanced research review methodology. However, the number of extant meta-analytic studies conducted to evaluate the effects of acceleration on high-ability learners is limited. A review of literature only identified four meta-analyses or best-evidence syntheses so far. They are the studies of Kulik and Kulik (1984), Rogers (1991), Kent (1992), and Kulik (2004). These existing meta-analytic studies have some limitations that need to be improved and overcome to obtain more accurate results. For instance, all of the four previous meta-analytic studies are heavily based on early studies of acceleration in the last century. The earliest study included was published in 1928 (see Rogers, 1991 for reference) and the most recent study included was published in 1991 (see Kent, 1992 for reference). It is noteworthy that the most recent original meta-analysis is Kent's (1992) study, since Kulik's (2004) meta-analysis actually was a re-analysis of the studies included in the three previous meta-analytic studies. This means the most recent studies in these meta-analyses were conducted more than 15 years ago. Educational practices and research have changed tremendously in the last decade. As Kulik (2004) pointed out, meta-analyses on the effects of acceleration were dated, since studies from recent years were not included in the analyses.

Furthermore, these existing four meta-analytic studies are not informative for education policy makers and other research consumers. First, these meta-analytic studies have different emphases. For instance, Kulik and Kulik's (1984) meta-analyses focused on the effects of acceleration on elementary and secondary school students; Rogers' (1991) best-evidence synthesis examined the effects of educational acceleration on gifted students; Kent's (1992) meta-analysis looked at the effects of acceleration on the social and emotional development of gifted elementary students; and Kulik' s (2004) meta-analysis investigated the academic, social and emotional effects of acceleration. In addition, a review of these meta-analytic studies reveals that even when they had very similar study foci (for example, all of the four meta-analytic studies examined social-emotional development), they actually included different sets of studies and reached different conclusions. Therefore, a new meta-analysis on the effects of acceleration is not only feasible, but also necessary.

In this study, the following research questions were addressed:

1. How does acceleration affect high ability learners' academic achievement?
2. How does acceleration affect high ability learners' social-emotional development?
3. Which conclusions from previous meta-analytic studies regarding the effects of acceleration are supported?
4. What differences exist between subject-based acceleration and grade-based acceleration in terms of their effects on high-ability learners?
5. What are the typical effects of the most common acceleration forms, such as grade skipping, early kindergarten/school/college entrance, advanced placement, subject-matter acceleration (e.g., math and reading), and curriculum compacting on high-ability learners?
6. What moderators are significantly associated with the effects of acceleration on high-ability learners?

1.2 Rationale

1.2.1 Assumptions

This meta-analysis was conducted under two basic assumptions. First, a strong research knowledge base exists on acceleration practices. Since the earliest acceleration practice appeared in 1862, broad forms of acceleration practice have evolved. Among the various acceleration forms, the most common ones include early school/college entrance (Brody & Stanley, 1991; Stanley, Keating, & Fox, 1974), subject-matter acceleration (Brody, Assouline, & Stanley, 1990), grade skipping (Daurio, 1979), curriculum compacting (Renzulli, Smith, & Reis, 1982), mentorship, Advanced Placement courses, academic summer programs (Lee & Olszewski-Kubilius, 2005; Southern & Jones, 1991; Southern, Jones, & Stanley, 1993), and self-paced classes (Lee & Olszewski-Kubilius, 2006). Based upon the extensive reviews of the experts on acceleration, the authors of *A Nation Deceived: How Schools Hold Back American's Brightest Students* (Colangelo et al., 2004) concluded that:

> Acceleration has been well researched and documented; Acceleration is the best educational intervention for high ability (gifted) students; Acceleration is consistently effective with gifted students. Acceleration is highly effective for academic achievement; acceleration is usually effective in terms of social-emotional adjustment (Volume II, p. 2)

Second, the effects of acceleration can be measured, which makes it possible to quantitatively synthesize the research findings. Extant literature shows that many research studies have used achievement scores, growth in individual subjects or all areas, self-esteem and academic self-concept, and attitudes toward school and learning, as probable ways to examine the effects of acceleration. Moreover, a number of previously completed meta-analyses or best-evidence syntheses regarding the effects of acceleration or ability grouping have demonstrated the possibility and promise of quantitatively synthesizing a large set of research findings. In summary, it can be concluded that the effects of acceleration are measurable, so it is possible to quantitatively synthesize the study outcomes.

Furthermore, this meta-analysis of the effects of acceleration corresponds to calls for "scientifically based research". According to the National Research Council (1999), evidence-based education is a comparatively new trend. The Education Sciences Reform Act of 2002 stated that education should be "transformed into an evidence-based field" (U. S. Department of Education, 2002, p.1). The National Research Council (NRC) report *Scientific Research in Education* has discussed the nature of education research and provided some standards for high-quality scientific inquiry in education (National Research Council, 2002a). After that, the National Research Council (2002b) report of *Advancing Scientific Research in Education* further offered 13 recommendations to promote actions for advancing scientific research in education. These 13 recommendations were organized around three strategic objectives: (a) promoting quality, (b) building the knowledge base, and (c) enhancing the professional development of researchers. According to this report, education researchers should especially pay attention to empirical inquires which emphasize the reanalysis, replication, and testing of existing theories, so education policy and practice can use what is known from research on a regular basis.

1.2.2 Advantages of Meta-Analysis

Meta-analysis has experienced impressive growth during the past two decades with widely-recognized advantages as a new research synthesis methodology. Glass (1976) introduced the term of meta-analysis as "…the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings" (p.3). He stated that meta-analysis connoted a rigorous alternative to the traditional narrative reviews of research studies, and it enabled researchers and reviewers to integrate and assimilate the vastly growing research literature.

One of the widely-recognized advantages of meta-analysis is that it enables researchers to overcome several problems of traditional literature reviews. Specifically, it is believed that meta-analysis methodology is superior to traditional research reviews in the following ways: (a) including studies based on specified inclusion criteria rather than on reviewers' subjective opinions, (b) providing more accurate interpretations of results, (c) examining study features as potential explanations, and (d) investigating the existence of moderating variables in the relationships being examined (Hedges & Olkin, 1985; Light &

Pillemer, 1984; Light & Smith, 1971; Wolf, 1986). Thompson (1999) concluded that meta-analysis could produce results with increased sample size, statistical power, and reliability of findings regarding treatment effects.

Many researchers have promoted the use of meta-analysis as a best way to extensively synthesize research findings. For example, Glass (1976) claimed that the development of the social sciences was hindered by heavy reliance on traditional narrative review methods, which were believed to be too informal and subjective. Therefore, he advocated the use of formal and quantitative methods in research reviews, which he called “meta-analysis”. Slavin (1986) pointed out that the advent of meta-analysis positively impacted research synthesis. Therefore, he proposed best-evidence synthesis as a thorough and unbiased means of synthesizing research, which actually combined the quantification and systematic literature search methods of meta-analysis with the detailed analysis of the traditional reviews. Likewise, Kulik (2004) noted that the traditional narrative review could put reviewers in a situation of being charged for bias and subjectivity since those methods failed to establish a clear relation between study findings and review conclusions, or failed to ensure comprehensive scrutiny of the literature. Furthermore, meta-analysis is believed to facilitate the effective translation of research findings into practice (Kavale, 1984; Kavale & Glass, 1984). More recently, meta-analysis has been considered an “indispensable” tool in current research synthesis, and “systematic reviews with meta-analyses occupy the top position in the hierarchy of evidence” (Bax, Yu, Ikeda, & Moons, 2007, p.6).

The existing literature suggests that meta-analysis has been recognized and widely used as an advanced research review methodology. For example, since the 1980s, there were at least five books published that describe the methods of meta-analyses (e.g., Glass, McGaw, & Smith, 1981; Hedges & Olkin, 1985; Hunter, Schmidt, & Jackson, 1982; Rosenthal, 1984; Wollf, 1986). Recently, there have been more books published that provide specialized instructions for conducting meta-analysis in particular fields, such as the field of health care and medical research (e.g., Egger, Davey, & Altman, 2001; Glasziou, Irwig, Bain, & Goldiz, 2001; Sutton, Abrams, Jones, Sheldon, & Song, 2000). As computer software has become a necessity for meta-analysis, numerous programs have been developed, such as the Comprehensive Meta-analysis (CMA) (Borenstein, Hedges, Higgins, & Rothstein, 2006), MetAnalysis (Leandro, 2005), MetaWin 2.1 (Rosenberg, Adams, & Gurevitch, 2000), and WEasyMA (Bax et al., 2007). According to Kulik and Kulik (1989), at least 100 meta-analyses of research findings had been conducted in the field of education. Furthermore, meta-analysis methods have been included in introductory research methods texts (Vockell & Asher, 1995). Generally, meta-analysis has been widely used in education, psychology, management, and the health sciences (Kulik, 2004; Pierce, 2007).

In the field of gifted education, researchers and educators have also provided a sound rationale to encourage the application of meta-analysis. For example, Asher (1986) noted that meta-analysis enabled educators and researchers to overcome interpretation obstacles such as imprecise measurement and the small sample sizes of gifted education research. Lauer and Asher (1988) defined meta-analysis as “a systematic, replicable, and relatively unbiased method of summarizing the overall results of a particular body of experimental research literature” (p.284). Vaughn, Feldhusen, and Asher (1991) stated that meta-analysis was a new and more comprehensive method to conduct program evaluations in gifted education. Again, Asher (2003) strongly encouraged meta-analysis in gifted education. He asserted that the generalization of the findings of a meta-analysis would be greater than any individual study; thus meta-analysis would provide the best evidence available to build bases for theory development.

1.2.3 Criticisms of Meta-Analysis

Publication bias. There is a growing body of research supporting the view that large and statistically significant effects are far more likely to be submitted to journals and published than those with non-significant findings. For example, according to Sterling (1959), 97% of the articles in psychology journals presented statistically significant findings. Greenwald (1975) demonstrated that statistically significant findings were eight times more likely to be submitted than non-significant ones. Similarly, White (1982) noted that unpublished studies with statistically significant findings often had significantly smaller effect sizes compared to the published ones.

A number of methods can be used to identify and correct for the publication bias problem. For example, Glass et al. (1981) suggested including unpublished papers, dissertations, reports and other unpublished sources in meta-analyses, to reduce publication bias. However, including these types of documents would raise a new issue… the qualitative differences between published and unpublished studies. It is commonly believed that most refereed journals have strict standards for publication, which imply that the studies that are published in those journals have more sound methodological characteristics and are more likely to provide evidence-based findings (Wang & Bushman, 1999). Eysenck (1978) pointed out that in cases where poor quality studies were included, a meta-analysis would end up as an exercise in “mega-silliness”, which only demonstrate the maxim “garbage in garbage out”.

Another commonly-used approach to correct publication bias problems involves determining the extent of the selection bias problem by estimating the number of unpublished null-result studies that would have to occupy file drawers in order to bring the estimate of the literature effect to the significance threshold (Rosenberg, 2005; Rosenthal, 1991). This approach is also known as the file drawer “fail-safe number” (Dooley, 2006). It provides a rough basis for judging the strength of the literature review conclusion against the alternative hypothesis that including the unpublished literature would nullify (Dooley, 2006). Dooley explained that a small fail-safe number would call into question the literature review effect because only a few negative findings would need to occupy file drawers to reduce the effect of the literature below significance, whereas a large fail-safe number implied the file drawer effect was not a serious problem.

In this meta-analysis, two approaches were used to detect and correct for publication bias. First, a funnel plot was produced and evaluated, as recommended by Light and Pillemer (1984). Second, Duvall and Tweedie's (2000) trim and fill procedure was conducted to further assess and adjust for publication bias. The Comprehensive Meta-Analysis (CMA) (Borenstein, Hedges, Higgins & Rothstein, 2006) software, with which this meta-analysis was conducted, made both of these approaches feasible. Moreover, the visual features of these two approaches are especially beneficial. The rationales, as well as the technical details of these two methods will be discussed in detail in Chapter 3.

The problem of finding integrity. As key proponents of the method of meta-analysis, Glass et al. (1981) pointed out that one of the main criticisms was that some meta-analyses resulted in aggregating findings from various research designs that were not commensurate in their main characteristics. In other words, it was trying to compare "apples and oranges". In addition, some researchers have also criticized meta-analyses for failing to maintain the contextual integrity of the individual studies that they were synthesizing. There are always some arguments between the authors of individual studies and reviewers who have conflicting opinions on various issues especially regarding the inclusion, or the interpretation of the results. Some researchers believe that all research syntheses integrate information from different studies (Cooper & Hedges, 1994; Glass et al., 1981), and so it is simply a matter of scope. Efforts are required to deal with the situations that occur when integrating studies that ask either very narrow or very broad research questions.

It has been suggested that coding a hierarchy of constructs can be helpful, in that the reviewers can begin with a more general question using broad constructs and proceed to more specific research questions with more specific constructs (Hedges & Olkin, 1985). In order to obtain a sufficient degree of finding integrity, Hedges and his colleague's above suggestions were adopted in this meta-analysis. Specifically, outcome variables or constructs that appeared in the primary studies were coded as comprehensively as possible (see Appendices D & E for a detailed description of the outcome variables coded). Moreover, the research questions that this meta-analysis addressed were constructed on the basis of some existing meta-analytic studies of the effects of acceleration on gifted students' academic achievement and/or social-emotional development.

The issue of dependence. Meta-analysis has been severely criticized for using "lumpy data" or combining non-independent results. Sipe and Curlette (1997) summarized five common sources of dependent data in meta-analysis: (a) more than one response from the same subject, (b) multiple measures from the same subject across time, (c) single outcome measure containing dependent scores (e.g., a subset and a global score), (d) dependent studies within a single document (same subjects used in two different studies reported in the same document), and (e) dependent samples across studies (same subjects in different studies). There are a number of methods recommended to address the issues of dependence. For example, Walberg and Haertel (1980) recommended a nested analysis of variance model. Glass et al. (1981) proposed using Tukey's jackknife procedure to deal with the interdependencies that arise when synthesizing a large set of research findings. Kulik and Kulik (1989) suggested using only one effect size from each study for any given analysis. However, since the middle of the 1990s, researchers have preferred the inclusion of all the effect sizes contained in the studies (e.g., Gleser & Olkin, 1994; Kalaian & Raudenbush, 1996; Lambert, 1995). It is noteworthy that although these approaches could be used to minimize violations regarding independence of effect sizes, they do not completely eliminate the problem of dependence (Cooper, 1998). In this meta-analysis the effect size dependence dilemma was handled by averaging the preliminary effect sizes within each primary study, as recommended by Hedges, Shymansky and Woodworth (1989). A brief description about this procedure will be provided in Chapter 3.

1.3 Conceptual and Operational Definition

Academic achievement. Academic achievement is one of the most important outcome variables in acceleration research. There is no single way to measure academic effects, so defining its indictors is a key to conducting a meta-analysis. In this meta-analysis, the definition of academic effects is based on a review of previous relevant meta-analytic studies and traditional narrative reviews. Rogers (2004) summarized the indicators of academic effects and the predictors of probable student success in terms of subject-based and grade-based acceleration. According to Rogers (2004), the following are indicators of probable student success:

- process and achievement well above age peers;
- mastery well above grade/age level in specific subject area or topic;
- two or more grade levels ahead;
- preference for fast-paced challenge and independence or small-group learning;
- having wide-ranging academic interests and being active in a variety of out-of-school activities;
- ability/achievement in upper 3%, and
- having strong achievement needs.

There are variations of measurements across different studies, which create difficulties for research synthesizers. Based on the existing research literature, outcomes in this study were categorized into P-12 level and post-secondary level academic achievement. For the P-12 level, the outcomes mainly included test results and the status of higher-education institutions to which accelerated high-ability learners were admitted. For post-secondary level, there were four subgroup outcomes:

1. Education background (e.g., degrees obtained, the status of higher-education institutions they attended)
2. College GPA
3. Ages when obtained certain degrees or reach any career achievement
4. Career status

Social-emotional development. Researchers and reviewers have used diverse terms to define the non-academic achievement related to effects of acceleration. For different terms, or even the identical terms, they refer to different constructs or variables. For instance, Kulik and Kulik (1984) grouped the effects of accelerated instruction into cognitive outcomes and non-cognitive outcomes. For evaluating the non-cognitive outcomes, they examined effects of acceleration on subject matter attitudes, attitudes toward school, vocational plans, participation in activities, popularity with peers, social adjustment, and character ratings. Rogers (1991) categorized non-academic outcomes of acceleration as socialization and psychological adjustment. In her study, the socialization construct includes the following variables: peer acceptance, self concept/confidence, social cognition, peer interaction, extracurricular participation, group conformity, leadership position held, career focus/satisfaction; the psychological adjustment construct covers the variables of emotional stability, referrals for counseling, mental maturity/adjustment, psychological traits (anxiety, creativity, etc.), conscientiousness, and rationality/irrationality.

Neihart (2007) stated that social effects were typically measured via "social maturity scores, teacher ratings of social skills, participation in extracurricular activities, and leadership positions held" (pp. 331-332); whereas emotional effects were usually examined by "measures of self-concept or teacher or parent ratings of risk taking, independence, and creativity" (p.332). Kent (1992) provided a most exhaustive list of variables regarding the social-emotional development (see Appendix A), including the indicators of the non-intellective aspects of participation, relationships, academic inclination, behavior, attitude, autonomy, analytical, and personality characteristics. The constructs and variables that appeared in existing studies described above served as a framework for this meta-analysis, especially for study coding.

The term "social-emotional development" is used in this meta-analysis, because the literature review suggests that this term has been most commonly used in the existing education and psychology literature (e.g., Kent, 1992; Kulik, 2004; Neihart, 2007; Robinson, 2004). Table 1 presents the variables that were chosen as indicators of social-emotional development in this meta-analysis.

Outcome measures are another factor that researchers need to consider when examining the effects of acceleration on social-emotional development. Self-reported questionnaires are very often used to measure social-emotional development. The existing literature shows that there are some psychologically sound personality inventories that have been used to investigate the social-emotional development among gifted and talented students. They include the Minnesota Multiphasic Personality Inventory (MMPI) (Tellegen, Ben-Porath, McNulty, Arbisi, Graham, & Kaemmer, 2003), and the California Psychological Inventory (Aiken, 2004).

Table 1. Study Variables Related to Social-Emotional Development

Construct	Variable	Description(sub-variables)
Social-development	Social relationship	e.g. peer acceptance/interaction, social cognition, group conformity
	Participation in extracurricular activities	
	Leadership position held	
Emotional-development	Psychological adjustment	e.g., emotional stability, referrals for counseling, mental maturity/adjustment, locus of control, flexibility, independence
	Level of satisfaction	e.g., satisfaction with life, career, choice of acceleration, etc.

Construct	Variable	Description(sub-variables)
Motivation	Self-image	e.g., self-concept, self-acceptance, self-reliance, self-esteem, self-confidence, etc.
	Educational plans	e.g., higher education aspiration
	Vocational plans	
	Liking school/subject/learning	e.g., attitude toward math, science, reading, etc.

In addition, to address the impact of acceleration on social-emotional development one should first consider the major prevailing concerns toward it. Robinson (2004) identified some sources of apprehension regarding the social and emotional effects of acceleration. She pointed out that when evaluating outcomes, it was necessary to remember the fact that acceleration was always considered as a solution to some existing problems, which implied students' dissatisfaction, unhappiness, loneliness, and loss of interests in school. She asserted that there were three major concerns pertaining to the social-emotional development effects of acceleration. They were: (a) the assumption that high-ability learners' emotional maturity should relate to their chronological age rather than mental age, (b) social issues, and (c) mental health issues.

High-ability learners. The subjects of the research studies included in this meta-analysis are high-ability learners. The term 'high-ability learners' primarily refers to academically gifted/talented students, who are defined as:

> Children and youth with outstanding talent perform or show the potential for performing at remarkably high levels of accomplishment when compared with others of their age, experience, or environment. Children and youth who exhibit high performance capability in intellectual, creative, and/or artistic areas, possess an unusual leadership capacity or excel in specific academic fields. They require activities not ordinarily provided in the school. Outstanding talents are present in children and youth from all cultural groups, across all economic strata, and in all areas of human endeavor (United Sates Department of Education, 1993, p.3).

This definition is preferred because it stresses the significance of potential and emphasizes comparisons between gifted students and others of equivalent age, experience, or environment. It also implies that giftedness exists in all cultural groups and across all socioeconomic levels. This definition is widely accepted in the field of gifted education.

The effects of acceleration on gifted students have been extensively studied. However, the term ‘high-ability learners’ is preferred and used in this study to avoid biased perceptions or opinions, which may subconsciously consider gifted students as those with impressive IQ, or extremely extraordinary achievement records. In addition, the use of high-ability learners reflects the fact that studies conducted and/or published outside the field of gifted education were included in the review.

Categorization of acceleration. There are two major approaches to categorizing acceleration forms. One is to define them as subject-based and grade-based acceleration (Rogers, 1992, 2004; Southern & Jones, 1991). Southern and Jones (1991) defined subject-based acceleration as instructions or practices that provided the learners with advanced knowledge, skills, and understandings in a particular content area or subject, before they reached a certain age or grade level. Rogers (1992) identified 14 forms of subject-based acceleration. These included: early entrance to kindergarten or first grade, compacted curriculum, single-subject acceleration, concurrent/dual enrollment, talent search programs, correspondence courses, distance learning, independent study, advanced placement courses, international baccalaureate programs, college-in-the-school programs, mentorship, credit for prior learning/testing out, and post-secondary options. Grade-based acceleration, according to Southern and Jones (1991), includes educational practices that enable learners to reduce the number of years they normally remain in the K-12 school system before their post-secondary education. According to Rogers (2002), there are five forms of grade-based acceleration: grade skipping, non-graded/multi-age classrooms, multi-grade/combination classrooms, grade telescoping, and early admission to college.

Another approach of categorization is to group acceleration interventions as radical acceleration and non-radical acceleration (Gross, 2004; Stanley, 1978). Radical acceleration, as defined by Stanley (1978), is instructions or strategies that enabled students to graduate from K-12 education three or more years earlier than usual. Gross (2004) stated that radical acceleration was particularly suited to young people who were “exceptionally” (IQ 160-179) or “profoundly” (IQ 180^+) gifted (p.87).

In this study, the following perspectives were used to analyze studies: (a) subject-based acceleration and grade-based acceleration, (b) acceleration with different grade levels (i.e., elementary, middle, secondary, and post-secondary levels), and (c) the most commonly implemented and studied acceleration forms, such as grade skipping, early entrance to school, early entrance to college (Neihart, 2007), Advanced Placement (AP), and international baccalaureate programs (Hertberg-Davis & Callahan, 2007). The reasons for these preferences were first, the categorization of subject-based and grade-based acceleration is widely used in acceleration research and the findings of a meta-analysis from this perspective will be very informative not only to educational researchers, but also to educational practitioners, and policy-makers. Second, these perspectives reflect the particular inquiry interests of the researcher. Specifically, the researcher was interested in exploring the effects of acceleration in terms of the grade levels, and particular forms of acceleration practices, especially those most commonly used and studied. Third, these perspectives provide important information for building subgroups when searching for possible moderators.

CHAPTER 2 LITERATURE REVIEW

Researchers have made concerted efforts to comprehensively examine the effects of acceleration, using traditional narrative reviews, meta-analytical methods, or a combination of both methods. The purpose of this chapter is to review and summarize the findings from recent narrative reviews of acceleration and tocritically review existing meta-analyses and best-evidence syntheses of the effects of acceleration. Accordingly, this chapter consists of three sections: (a) critical review of recent traditional narrative reviews, (b) critical review of meta-analyses and best-evidence syntheses, and (c) conclusions from the literature review.

2.1 Review of Recent Narrative Reviews of Acceleration

The review of traditional narrative reviews of acceleration in this chapter is limited to those conducted since 1984, when the first meta-analysis of the effects of acceleration was published (i.e., Kulik & Kulik, 1984) to present. The reviews included either represent the most recent collective scholarly work regarding the issue of acceleration, or exemplify the newest comprehensive reviews by some influential experts on acceleration.

Acceleration has been viewed as one of the most important issues in gifted education and many researchers have devoted decades to studying it. A number of collective scholarly works have been produced through synergized efforts. These collective works best represent the expertise and research findings from both the field of gifted education and general education; therefore they provide a solid evidence base for acceleration practice and research. For example, the Institute for Research and Policy on Acceleration (IRPA), under the sponsorship of The Connie Belin & Jacqueline N. Blank International Center for Gifted

Education and Talent Development (Belin-Blank Center), at the University of Iowa, is one of the most well-known institutions focusing on promoting practices and research on acceleration. Attributed to the efforts of the Belin-Blank Center, as well as experts from the field of gifted education and general education, *A Nation Deceived: How Schools Hold Back America's Brightest Students* (Colangelo et al., 2004), represents the most recent and comprehensive collective work regarding the issue of acceleration. In addition, some influential researchers and educators, including Nicholas Colangelo, Miraca Gross, David Lubinski, Nancy Robinson, Karen Rogers, and Julian Stanley have focused their research on acceleration over several decades. They have been systematically updating and synthesizing their research over the years. Their most recent work, especially their comprehensive reviews, usually covers key research findings from their decades of work, and represent years of accumulated knowledge, expertise, and insights.

The Templeton National Report on Acceleration. The Templeton Report on Acceleration (Colangelo et al., 2004) is a national report on acceleration. It was produced by the collective efforts of distinguished researchers and educators on acceleration from all over the world. The purpose of this report was to provide all the existing research findings and information that American educators should know, in order to make evidence-based decisions on educating high-ability learners. This was necessary because many parents, teachers, and administrators had not accepted acceleration as a common practice because they were not convinced of its effectiveness. The report is comprised of two volumes. Volume II presents an extensive review of the research on acceleration of high-ability learners, and Volume I articulates the most important points from the research reviews presented in Volume II. Table 2 summarizes the 20 most important points from this report that represent the overall research findings about acceleration.

Table 2. Summary of Key Points of the Templeton National Report on Acceleration

Overall effects
• Acceleration is the most effective intervention option for high-ability learners. • Acceleration is economical.
Academic, social-emotional effects of acceleration
• Acceleration has long-term positive effects on high-ability learners, academically and socially. • Acceleration provides better social and emotional cues for high-ability learners. • The existing 18 types of acceleration can be categorized into grade-based acceleration and subject-based acceleration. The former could allow high-ability learners to reduce the number of years in the K–12 education, whereas the latter provides them the opportunities for advanced content earlier than regular paths. • Early school entrance is a good option for high-ability learners, in light of academic and social effects. • Early college entrants are more likely to experience both short-term and long-term academic success, long term career success and personal happiness. • It is rare that early college entrants have social or emotional difficulties. • Radical acceleration is particularly effective for highly gifted students both academically and socially.
The necessity of acceleration
• High-ability learners would suffer from boredom and lose learning interests if they were only provided the curriculum which serves for their age-peers. • Educational equity respects individual learning differences and values every student.

Accessibility of acceleration

- There are evidences and systems to facilitate schools to make informed decisions about acceleration.
- Aside from the early college entrance, there are many alternatives for high-ability high school students, such as Advanced Placement (AP), dual enrollment in high school and college, distance education, and summer programs.
- Testing, especially that developed for high-ability learners’ older-age peers, is highly effective in identifying students for acceleration.

System for facilitating acceleration

- It is necessary to rely on legislation, the courts, administrative rules, and professional initiatives, in order to ensure a significant change in the people’s perceptions toward acceleration.
- High-ability learners with disabilities need more time and resources to ensure the implementation of acceleration.
- Parents of high-ability learners need to facilitate their children’s decision-making on acceleration.

What needs to be improved

- Some problems arising from acceleration practices are actually attribute to incomplete or poor planning.
- Educators commonly perceive the acceleration practices negatively, although researchers have provided numerous evidence regarding the benefits of acceleration.
- The question that educators should ask is how to better accelerate high-ability learners.

This report also revealed six pieces of evidence to support the point that schools don't serve high-ability learners well in acceleration implementation. These include:

1. School administrators, educators, parents and students still lack familiarity with the research findings on acceleration.
2. There exists a preoccupied perception that children should always stay with their age group.
3. It is believed that children would be deprived of their precious childhood because of acceleration.
4. People are afraid that acceleration harms children socially.
5. People are concerned about the issue of educational equity.
6. Some people are concerned that other students would be treated differently if acceleration opportunities are only open to some children.

Reviews by influential experts on acceleration. As an expert who has consistently contributed to research on acceleration, Karen Rogers has focused on acceleration and conducted several reviews (e.g., Rogers, 1991, 1992, 2002, 2004, 2007). Rogers' (2002) synthesis suggested that gifted students benefited from advanced placement or international baccalaureate by about one third of a year's additional academic growth, and they obtained three fifths of a year's academic gain from mentorships and subject acceleration. She also noted, that both grade-based and subject-based acceleration were beneficial to children's academic achievement, as long as they were self-directed, motivated, and willing to work beyond their grade levels.

Rogers (2004) concluded that "the academic outcomes of acceleration are impressive" (p. 56) and there was a large body of research supporting acceleration options for gifted learners in the field of gifted education. She further insisted that in order to decide the specific forms of acceleration for gifted learners, it was necessary to know the general academic effects of acceleration. Several of her remarks are especially noteworthy. For example, she pointed out that "individual student readiness is critical" (p.56), therefore she urged that educators and schools should collect enough supplementary information regarding an individual student's cognitive characteristics, learning strengths, learning preferences, and interest and activities. Schools should consider which forms of acceleration would be most appropriate and effective to meet students' academic and social-emotional needs.

Swiatek (2007) examined the development of the Talent Search Model and reviewed the effects of Talent Search educational programs. She concluded that research has shown that both academic and psychological, and short-term/long-term benefits to the participants of talent search programs, which offer the high-ability learners various educational opportunities, such as enrichment-based residential programs, and summer or weekend accelerative classes. For example, she found that that high-ability learners who attended the first Study of Mathematically Precocious Youth (SMPY) fast-paced math class exhibited higher achievement in high school, were more likely to choose advanced mathematics classes than others, were more willing to accept acceleration, attended more prestigious colleges, and preferred graduate studies.

Having studied radical acceleration over decades, Gross (2004) listed eight predictors of successful radical acceleration, which include: early acceleration, student engagement in educational planning, family support, informative and supportive mentors, a broad range of acceleration opportunities, above average performance, exposure to advanced content prior to acceleration, and self-knowledge and careful pre-planning. Gross concluded that high-ability accelerants generally experienced very high levels of success in higher education, and they also entered in high-status professions. In summary, Gross noted that radical acceleration was a practical and cost-efficient educational intervention for high-ability learners.

Reviewing the research on radical acceleration and early college entrance, Gross and Vliet (2005) asserted that the academic effects of radical acceleration on gifted students could be considered "highly impressive" (p.168). They stated that, as documented in numerous studies, impressive academic achievements are indicated by: earning higher GPAs, having a high probability of finishing college on time or even early, earning honors, making the dean's list, pursuing graduate degrees, engaging in research, and embarking on prestigious careers (Olszewski-Kubilius, 1995; Stanley, 1978; Swiatek & Benbow, 1991; Terman & Oden, 1959). Regarding the social-emotional effects, they stated that research suggested positive social and emotional outcomes, which was usually supported by the evidence that students who had been radically accelerated adjusted well to new learning contexts, made friends without difficulty, were accepted by older students, and experienced increased levels of self-esteem and self-confidence (Gross, 2003; Pollins, 1983). Gross and Vliet (2005) concluded that:

> Research provides strong support for the use of thoughtfully planned and monitored radical acceleration as a process allowing educators to respond to the academic and affective needs of a significant subgroup of the gifted population (p. 168)

Southern and Jones (2004) identified five dimensions of acceleration: pacing, salience, peers, access, and timing, which can characterize and affect the availability of acceleration to students who are qualified for the opportunities. They discussed issues in acceleration practices, such as the unintended consequences from acceleration, the difficult decision of pacing and curricula, and the unavoidable interaction with bureaucratic entities. They concluded that acceleration serves a variety of purposes, such as enabling students to learn more quickly with less help from teachers, and allowing students to finish standard curriculum earlier than age-/grade-level peers, therefore a broad range of acceleration options are needed.

Lubinski (2004) reviewed four key longitudinal studies of the Study of Mathematically Precocious Youth (SMPY), to examine the long-term effects of educational acceleration. These four studies are: Study 1, Lubinski, Webb, Morelock and Benbow' s (2001) 10-year study of the top 1 in 10,000 in mathematical or verbal reasoning; Study 2, Benbow, Lubinski, Shea, and

Eftekhari-Sanjani’ s (2000) 20-year longitudinal study of the top 1% in mathematical reasoning ability; Study 3, Bleske-Rechek, Lubinski, and Benbow’ s (2004) study of three decades of longitudinal data on the advanced placement (AP) programs; Study 4, Lubinski, Bendow, Shea, Eftekhari-Sanjani, and Halvorson’ s (2001) 20-year comparison study of top math/science graduate students with same-age SMPY participants. Overall these researchers have found that: (a) academically precocious students who took advantage of acceleration opportunities (i.e., Talent Search program) viewed their experiences positively when they grew up, (b) academically precocious students who experienced acceleration during middle schools and high schools rated their pre-college learning experiences much higher than their peers of the equivalent aptitudes who didn’t have acceleration experiences, and (c) acceleration experiences seem to be very important for nurturing high level scientific leaders. The authors also found that there was great psychological diversity within academically precocious students across both cognitive and noncognitive attributes relevant to educational and occupational contexts. In addition, they found social-emotional factors, such as interests, values, and time willing to study or work, had been relatively neglected, although they were critical for the research and practices of acceleration.

Colangelo, Assouline, and Lupkowski-Shoplik (2004) stated that there were systems that could ensure whole-grade acceleration would be be a “low-risk/high-success intervention” (p. 85) for qualified high-ability learners. They asserted that the *Iowa Acceleration Scale* (IAS) (Assouline, Colangelo, Lupkowski-Shoplik, Lipscomb, & Forstadt, 2003) was a systematic and viable instrument to provide educators and parents with information and instructions for decision-making about whole-grade acceleration. The IAS can measure five salient factors that need to be considered when making a decision about acceleration for high-ability learners. These factors include: (a) academic ability, aptitude, and achievement; (b) school and academic factors; (c) developmental factors; (d) interpersonal skills; and (e) attitude and support. They stated that research on early entrance to kindergarten and first grade had shown that these experiences had a positive influence on the accelerants. They identified six characteristics of high-ability preschoolers which were demonstrated in previous research. Those characteristics include “early verbal ability … and early reading”, “strong mathematical skills”, “long attention span”, “extraordinary memory”, “abstract reasoning ability… and make connections between areas of learning”, and “an early interest in time” (p.83). This anecdotal information, mainly from parents, can be used to identify high-ability young children for acceleration.

Reviewing the effects of academic acceleration on the social-emotional status of high-ability learners, Robinson (2004) concluded that:

> The overwhelming evidence suggests that all the forms of academic acceleration constitute viable options as part of any attempt to provide an optimal educational and social match for high-ability learners. None of the options has been shown to do psychosocial damage to gifted students as a group; when effects are noted, they are usually (but not invariably) in a positive direction (p. 64)

Brody, Muratori, and Stanley (2004) summarized research findings to counsel high-ability learners who consider early college entrance as an option for their particular educational needs. They found that research on the academic adjustment of early college entrants provided much credence to early entrance to college as a successful intervention to meeting the needs of high-ability learners. They cited the Study of Mathematically Precocious Youth (SMPY), the Early Entrance Program at the University of Washington, and the National Academy of Arts, Sciences, and Engineering (NAASE) of the University of Iowa as exemplary early college entrance programs. The students who have entered these three programs have been extensively studied, and collectively show high-achievement. Conversely, there is also evidence that in some cases, early college entrants experience difficulty and fail to meet their own expectations. Brody et al. accordingly offered nine recommendations to guide high-ability learners when considering early college entrance. They suggested that high-ability learners:

1. Should take challenges in high school by participating in local, regional and national competitions.
2. May choose to complete some college-level courses, such as AP, 12th-grade honors, or part-time college courses, before they officially enroll as full-time college students.
3. Can try to interact with older peers through higher level classes or summer programs when they are still in high school.
4. Should consider that for early college entrance they should meet the criteria that their SAT I (aptitude) and SAT II (achievement) scores should be at least average or above average when compared to those who actually attend the particular university.
5. Should know if their language, critical reading, mathematics, computer, and study skills would enable them to succeed in college.
6. Should be highly motivated.
7. Should be willing to give up some extracurricular activities in high school.
8. Should be satisfied with the college they will attend.
9. Should avoid publicity, for some early college entrants (p.105).

Moon and Reis (2004) reviewed and synthesized the research on twice-exceptional students. Their review revealed that some characteristics of high ability students with learning disabilities (GT/LD) may disguise their need for acceleration and bring more challenges to increase the accessibility of acceleration. These characteristics include high levels of creative potential, disruptive or withdrawn behavior, underachievement, emotional intensity, unrealistic self-expectations, feelings of learned helplessness and low self-esteem (e.g., Baum & Owen, 1988; Dabrowski & Piechowski, 1977; Reis, Neu, & McGuire, 1995). In this review they also found that several forms of acceleration had been suggested for implementation with GT/LD students, such as acceleration (Olenchak & Reis, 2001), grade acceleration (McCoach, Kehle, Bray, & Siegle, 2001), curriculum compacting (Reis, Burns, & Renzulli, 1992), and mentorships (Olenchak, 1994, 1995). They concluded that researchers had generally agreed with the point that twice-exceptional students could benefit from acceleration, especially when acceleration intervention were implemented in a way that matched their interests, and in a challenging and supportive environment.

Olszewski-Kubilius (2004) provided an overview of Talent Search programs and accelerated programming for high-ability learners. According to this review, research suggests that Talent Search scores have been considered as a valid indicator of reasoning and learning rate, and there is evidence that acceleration is an appropriate and successful educational intervention for high-ability learners who are selected based upon talent search scores. This review concluded that the acceleration experiences related to the Talent Search programs show many positive influences, and they positively affect accelerants' high school and college education and academic self-efficacy. For example, this review revealed that there was evidence showing that students who participated in a kind of Talent Search programs were more likely to be admitted to more selective high education institutions, and to begin undergraduate study early. Olszewski-Kubilius therefore asserted that the Talent Search Model can be considered as one of the most outstanding forms of acceleration in gifted education.

2.2 Meta-Analytic Studies of the Effects of Acceleration

Education literature revealed that three meta-analyses of the effects of acceleration have been conducted (i.e., Kent, 1992; Kulik & Kulik, 1984; Kulik, 2004). In addition, Rogers (1991) also investigated the effects of acceleration, utilizing Slavin's (1986) "best-evidence synthesis", a comprehensive review method of incorporating the best features of both meta-analysis and traditional review. Furthermore, there have been a number of meta-analyses of the effects of ability grouping, among which the issue of acceleration has been partially addressed (e. g., Kulik, 1991; Kulik & Kulik, 1991; Slavin, 1987, 1990, 1993). The following section will first present the summary of the three previous meta-analyses and one best-evidence synthesis of the effects of acceleration (see Table 3). Following that, findings regarding the effects of acceleration from other relevant meta-analytic reviews of ability grouping will be reported.

Table 3. Summary of Four Previous Meta-Analytic Studies of Acceleration

	Kulik & Kulik, 1984	Rogers, 1991	Kent, 1992	Kulik, 2004
Study focus	Achievement Other	Academic Socialization Psychological	Social & emotional	Academic, social & emotional
Grades	Elementary Secondary	All levels	Elementary	All levels
No. of studies	26	247, only 81 studies with ES obtainable	23	25
Study year	1932-1974	1928-1987	1928-1987	1932-1986

	Kulik & Kulik, 1984	Rogers, 1991	Kent, 1992	Kulik, 2004
Interpretation criteria	Cohen’s (1977)	Suggestion from the National Institute of Education’s Joint Dissemination Review Panel	Modified Cohen’s (1977)	Cohen’s (1977)

	Kulik & Kulik, 1984	Rogers, 1991	Kent, 1992	Kulik, 2004
Findings	Accelerated vs. non-accelerated same-age peers: median ES of .80. Accelerated vs. non-accelerated older peers: median ES of .04.	Academic: Mean ES of .57. Socialization: mean ES of .17. Psychological: mean ES of .17.	Overall effect size: .13. Telescoping showed the greatest ES (.15). Kindergarten accelerants had the most positive ES (.14). Boys (.21) gained more than girls (.15).	Academic: Median ES of .80, with same-age control group. Median ES of -.04, with older control groups. Social-emotional: Ed. plans: positive influences. Attitudes: inconsistent effects. Activities: very little effects. Self-acceptance and personal adjustment: trivial effects

	Kulik & Kulik, 1984	Rogers, 1991	Kent, 1992	Kulik, 2004
Conclusion	Significant benefits in academic achievement.	Academic effects are significant; Overall socialization outcomes are significant.	No harmful effects on the social and emotional development of gifted elementary students	The academic effects are practically significant.

2.3 Other Relevant Meta-Analytic Studies

Slavin (1987) conducted a “best-evidence synthesis” to investigate the achievement effects of ability grouping in elementary schools. He concluded that research provided strong evidence for the positive effects of the Joplin Plan (cross-grade ability grouping for reading only) with a median effect size of .45. In addition, this study suggested that within-class ability grouping in the subject of mathematics appeared to be effective with a median effect size of .34. However, overall, the median effect size was .00, suggesting that there were no significant effects of ability grouping in elementary school students.

Slavin (1990) examined the achievement effects of ability grouping on secondary school students, using the method of best-evidence synthesis. He found that the overall effects of ability grouping on student achievement were close to zero at all grade levels, regardless of students’ prior performance, although it did appear that students in grades 7-9 were affected more by the grouping experience than students in grades 10-12. In addition, he claimed that ability grouping in subjects was ineffective in improving student achievement, and ability grouping of social studies actually showed negative influences. The findings from this study are inconsistent with those from some previous studies, which, have generally concluded that higher achievers benefit from ability grouping while low achievers are negatively influenced by the experiences of ability grouping.

However, there are some limitations in this study. First, as Slavin (1990) pointed out, none of the studies reviewed provided strong evidence to simply attribute the outcomes to the ability grouping practices. In other words, it is quite possible that there are some alternative explanations for the outcomes. For instance, the change of student achievement may relate to the differences of teacher behaviors or classroom characteristics. Second, almost all studies analyzed heavily relied on standardized tests as the measure of the outcomes, which may lead to incomplete measurement or missing information. Therefore, there is a need to consider including studies which used broader and more sensitive achievement measures for future research synthesis.

Kulik and Kulik (1991) conducted a meta-analytic review of the effects of grouping programs on student achievement. According to this review, programs of enrichment and acceleration, with the greatest amount of curricular adjustment, have the largest positive effects on student achievement. They concluded that the findings did not provide evidence to the then-prevailing views that no one benefited from grouping or that those in the lower groups were harmed academically or emotionally by grouping.

Vaughn, Feldhusen and Asher (1991) evaluated the effectiveness of pull-out programs in gifted education, combining meta-analysis and narrative review. This study indicated that pull-out models positively affected the achievement, critical thinking, and creativity of gifted students, with the effect sizes of .65, .44 and .32, respectively. However, they found that there was no evidence to support the view that pull-out programs influenced gifted students' self-concepts. It is noteworthy that this meta-analysis was only based on nine experimental studies. In fact, the effect sizes were calculated from synthesizing only 2 or 3 studies for each variable. Moreover, effects sizes were computed with different methods, due to the variations of the data reporting of the included studies. These facts should be considered when interpreting the results.

Slavin (1993) further synthesized research on the effects of ability grouping on the achievement of middle grade students. In fact, in this synthesis he extracted some studies which were originally included in his (1990) review, and also added studies that have appeared since 1990. This review showed that for middle and junior high school grades (6-9) students, overall achievement effects of ability grouping were close to 0, whether prior performance levels were high, average, or low. This review concluded that there was no evidence for the presumption that middle grade high achievers benefited more from grouping experience while low achievers were negatively affected by the grouping; the effects of between-class ability grouping on academic achievement were close near zero for students at all learning aptitudes.

2.4 Conclusions from the Literature Review

First, an updated meta-analysis on the effects of acceleration is warranted. There were only a limited number of meta-analyses evaluating the effects of acceleration, although there were a large number of traditional narrative reviews addressing this issue. The authors of *A Nation Deceived: How Schools Hold Back America's Brightest Students* (Colangelo et al., 2004) identified six publications of meta-analytic studies investigating the effects of acceleration, which they included into the annotated bibliography of the report. However, a further examination of the six publications revealed that only three of them were original meta-analysis or best-evidence synthesis studies (i.e., Kent, 1992; Kulik & Kulik, 1984; Rogers, 1991), which directly addressed the issue of acceleration. Among the remaining three studies, Kulik's (1993) study is an analysis of the research on ability grouping, Kulik and Kulik's (1992) study is a brief report of their meta-analysis of ability grouping in 1993.

Furthermore, although Kulik's (2004) study can be counted as an original meta-analysis in the literature review above, it is noteworthy that this study is in fact a re-analysis of the other three existing meta-analyses/best-evidence synthesis regarding acceleration. In other words, it was conducted by making use of the studies that had been included in the prior meta-analytic research of Kulik and Kulik (1984), Rogers (1991) and Kent (1992). The author did not conduct a new literature search to include any new research studies into the analysis. In fact, the newest study included in this meta-analysis was published in 1991 (i.e., Cornell, Callahan, & Loyd, 1991), which was originally included in Kent's (1992) meta-analysis of the social and emotional effects of acceleration. It is possible that schools, students, and acceleration practices have changed so much that the findings based on these old studies would not be able to provide much meaningful information or insights for today's education practices and research.

Therefore, it is necessary to conduct a new meta-analysis with an exhaustive search of the education literature so new research studies can be included.

Second, previous meta-analytic studies of the effects of acceleration generally only included a small number of studies for review. For example, the well-known meta-analysis by Kulik and Kulik (1984) included 26 controlled studies on the effects of acceleration on elementary and secondary school students. Among the 26 studies included, most of them are fairly old, with publication dates ranging from 1932 to 1970. After grouping them into studies with same age control groups and those with older age control groups, each group only has 13 studies. Kent (1992) included 23 studies for the meta-analysis of the effects of acceleration on the social and emotional development of high-ability elementary students. Furthermore, among the 23 studies included, some did not have appropriate comparison groups. For example, according to Kent (1992)'s summary of the included 23 studies, Study 1 (i.e., Arends & Ford, 1964) used 146 academically talented students as the experimental group and 58 regular students as the control group, which is not appropriate. Kulik (2004) pulled 25 studies from previous meta-analytic research for his analysis of effects of acceleration. Although Rogers' (1991) best-evidence synthesis included 247 studies, only 81 of the studies provided information from which effect sizes could be calculated. In addition, among the studies included some are not primarily relevant to acceleration. Finally, a new meta-analysis of the effects of acceleration can provide more research-based evidence for educational policy-making. This notion is mainly based on the features of the recent education literature. For example, the literature review indicates that a number of longitudinal research studies (e.g., Benbow et al., 2000; Bleske-Rechek et al., 2004; Lubinski et al., 2001) have appeared since 1984. These studies provided important information regarding the long-term effects of acceleration. A new meta-analysis, with the inclusion of these longitudinal studies, will provide information that was not available in the previous meta-analytic studies. Similarly, the recent emphasis of the Jacob K. Javits Gifted and Talented Students

Education program (U.S. Department of Education, 2008) on intervention research has stimulated new studies of accelerative option. In addition, one can expect that it is possible to draw a clearer picture about the social-emotional effects of acceleration on the basis of new studies focused on social-emotional issues. Previously, some meta-analysis researchers have claimed that a shortage of studies addressing the social-emotional effects of acceleration existed; as a result it is hard to draw any conclusions. The literature above indicated that there are more and more studies examining the social-emotional effects of acceleration, corresponding to the concerns and the calls for efforts regarding this issue. The purpose of this meta-analysis is to synthesize the results of all of these new studies, in order to provide a more comprehensive picture of the academic and social/emotional effects of acceleration.

CHAPTER 3 METHODS

This chapter describes the methods with which this meta-analysis was conducted. Information is provided about methods of searching the literature to locate relevant studies of acceleration, study inclusion criteria, study coding, study quality evaluation, effect size calculation and extraction, combination and grouping, data analysis, as well as the procedures of publication bias assessment. In addition to describing a given method, relevant justifications for the usage of each method are discussed, as needed.

3.1 Overview of the Method of Meta-Analysis

Meta-analysis is a research review methodology that uses common scales to synthesize existing quantitative research findings (Glass, McGaw, & Smith, 1981; Hedges & Olkin, 1985; Hedges, Shymansky, & Woodworth, 1989; Slavin, 1986). Meta-analysis enables researchers to investigate a broad variety of questions and to obtain the highest level of evidence, as long as a reasonable body of primary research studies exists. The basic procedures for a meta-analysis include: first, locating studies on a particular issue through objective and replicable searches; second, coding the studies for major features; and third, describing outcomes using a common scale. Meta-analyses have been widely used to summarize research findings in education, psychology, management, and health sciences (Kulik, 2004; Pierce, 2007).

Two types of meta-analysis have emerged in research literature. One type refers to the combination of hypothesis tests, whereas the other type entails the combination of estimates of treatment effects (Hedges, 1992). This meta-analysis follows the tradition of synthesizing estimates of treatment effects, which typically involves combining research studies with similar designs and outcome measures on a construct of interest. The preference for this tradition is based on its advantage of being able to provide information about the magnitude of treatment effects, which is not possible in meta-analyses using the combination of hypothesis tests. The first procedure in the estimation of treatment effects involves the selection of an index of effect magnitude, which can appear in various forms such as raw means, standardized differences between treatment and comparison groups, correlation coefficients, or the odds ratios between treatment and comparison groups (Hedges, 1992). In this study, the term “meta-analysis” refers to the meta-analysis tradition of the combination of estimates of treatment effects.

In this study, effect size was used as the common scale to evaluate the effects of acceleration on high-ability learners. Specifically, this meta-analysis was conducted using the five procedures described by Cooper and Hedges (1994): (a) problem formulation, (b) literature search, (c) coding studies, (d) calculating effect sizes, and (e) analyzing the data and interpreting the results. However, to ensure the soundness of the conclusions, an evaluation of study quality and assessment of publication bias was added.

3.2 Problem Formulation

Problem formulation involves the following three procedures: (a) define (conceptually and operationally) the variables of research interest; (b) identify the type of relationships of research interest; and (c) provide the historical, theoretical, and/or practical background of the research problem (Cooper, 2007). It can be concluded that the purpose of these procedures is to help the researcher to identify and clarify the operational concepts and variables involved. Simply put, these procedures are to establish a landscape for the review. As Cooper (2007) noted, this is a way to ensure the concepts and operations are neither too narrow nor too broad, therefore allowing an appropriate interpretation of results. For this meta-analysis, the problem formulation and the definitions of key concepts and variables were presented in the Chapter 1, whereas the rest of the method procedures are described in this chapter.

3.3 Literature Search

Studies conducted during the years from 1984 to 2008 were the focus of this meta-analysis. The year 1984 was chosen as a cut-off time because the first meta-analysis of the effects of acceleration was published in 1984 (i.e., Kulik & Kulik, 1984). Multistage procedures were conducted to search and identify studies for inclusion: (a) a search of electronic databases including Education Resources Information Center (ERIC), Web of Science, Psychological Abstracts Index (PsycINFO), ProQuest Dissertation & Theses, (b) web searches using Google and Google Scholar search engines , (c) a hand search through three primary empirical journals in gifted education: *Gifted Child Quarterly, Roeper Review*, and *Journal for the Education of the Gifted*, (d) a manual scrutiny of reference/bibliography lists of the other relevant previous systematic reviews and meta-analyses to locate additional relevant studies, and (e) personal contact with researchers in the gifted education field, to solicit ongoing research.

These searches were guided by a conceptual definition of acceleration and some frequently used terms for acceleration practice as identified in the Templeton National Report on Acceleration (Colangelo, Assouline, & Gross, 2004), as well as the Thesaurus of ERIC Descriptors and the Thesaurus of Psychological Index Terms. Therefore, the search keywords for the “acceleration” feature of research studies included: acceleration/ accelerated/accelerants/accelerating, early entrance, early admission, grade-skipping, continuous progress, self-paced progress, combined classes, curriculum compacting, telescoping, mentoring, extracurricular programs, correspondence course, early graduation, concurrent/dual enrollment, advanced placement, credit by examination, academic summer programs, fast-paced classes, fast-track, flexible progression. The search keywords for the “high-ability learners” features of the research studies included: gifted/talented, high ability, high achievement/achievers, accelerated/able learner, accelerants, high IQ, superior learners/students, advanced learners/students, advanced courses/programs. Studies of ability grouping and enrichment programs were also reviewed because they were often associated with accelerative practices.

3.4 Study Inclusion Criteria

To specify what kinds of studies were included or excluded in this meta-analysis, study inclusion criteria were formulated in light of the key research questions. The criteria covered the issues of study designs, populations, interventions, comparisons, and outcome measures, which are closely related to the eligibilities of the studies included.

1. *Scholarship*: Studies had to be empirical in nature. Publication venues included both peer-reviewed journal articles and other unpublished research documents such as dissertations, conference presentations, or research technical reports.
2. *Relevance*: The target subjects of acceleration described in each study had to be high-ability learners and the identification procedures for the high-ability participants must have been specified. Studies outside the fields of gifted education were included, as long as they dealt with acceleration of high-ability learners and met other inclusion criteria.
3. *Evident study design*: Studies had to have a recognizable study design that yielded outcomes suitable for meta-analysis, such as an experimental, a quasi-experimental, or a causal-comparative design.
4. *Measured outcomes*: At least one of the two major effects of acceleration on high-ability learners had to be reported in the studies: academic achievement, and/or social-emotional development, both for the accelerated group and the comparison group.
5. *Enough quantitative information*: Studies had to report credible, quantitative information, either in descriptive or inferential forms (e.g., means, standard deviation, t-values, F-values, ANOVA tables, or p-values), so that effect sizes could be calculated or estimated.
6. *Appropriate comparison groups:* The comparison groups and the accelerated groups had to be matched in terms of subjects' major aptitudes. Appropriate

comparisons included, for example, accelerated high-ability learners versus non-accelerated high-ability learners, or accelerated high-ability learners versus non-accelerated ones (oftentimes older age peers) in the equivalent grade level. Studies using inappropriate comparison groups were excluded. For example, according to Kulik and Kulik (1984), comparisons between talented and average youngsters seldom yielded useful information about program effects. There was one exception, however: when inappropriate comparison group was used but appropriate statistical control was applied so that the study results were obtained after the difference from nonequivalent learning aptitude was taken into account, these studies were still qualified to be included.

7. *Completeness and non-redundancies*: To avoid a situation of a single study exerting a disproportionate influence on the overall results, if the same study was reported in several papers, only the most complete publication was included for analysis. However, when an individual paper reported results from several studies only those results that were relevant and met all the inclusion criteria were included.
8. *English*: Studies had to be available in English. Studies from countries outside of the U.S. (i.e., Australia, UK, Spain, and China) but presented in English were also included.

3.5 Study Coding

Study coding was conducted with a coding protocol (see Appendix B *the Study Substantive Features Coding Descriptors*). This coding protocol was created with a series of successive efforts. First, a draft of the coding sheet was critiqued by an expert in the field of gifted education and revised several times. Pilot coding practices were then performed with six studies, and this led to further revisions. Furthermore, a review and discussion of the revised coding sheet was conducted with an expert on meta-analysis. Study coding was completed by the author of this meta-analysis. Independent double coding of two (5.3%) randomly selected studies was conducted by another graduate student at Purdue

University who had meta-analysis research experience and the author of this meta-analysis.

3.6 Study Quality Evaluation

As Wortman (1994) argued, judgments about the methodological quality of research studies are based on four criteria: internal validity, external validity, statistical conclusion validity, and construct validity. One common approach for evaluating study quality uses the “threats to validity” approach developed by Campbell and his colleagues (Campbell & Stanley, 1966; Cook & Campbell, 1979; Shadish, Cook, & Campbell, 2002). In this study, the approach of evaluating study qualities is based on Cook and Campbell’s (1979) definitions of the validities. Specifically, internal validity refers to the extent to which statements can be made regarding a causal relationship between one variable and another; external validity is the accuracy with which the assumed causal relationship can be generalized across various subjects, settings, or times; statistical conclusion validity denotes the accuracy of a statement about covariation between presumed independent and dependent variables; and construct validity is defined as the accuracy with which generalizations can be made about higher-order constructs from research operations.

Farrington (2003) proposed a fifth criterion, descriptive validity, which is defined as the extent to which major study findings and results can be presented appropriately in research reports. Farrington stressed that there was a need to develop standard measures to evaluate study validity. He further proposed that when designing a quality scale, the five validity types should be prioritized as follows: (a) internal validity, (b) descriptive validity, (c) statistical conclusion validity, (d) construct validity, and (e) external validity. Furthermore, Farrington (2003) recommended using a validity measure scale for each validity type: 0 (very poor), 1 (poor), 2 (adequate), 3 (good), or 4 (very good).

Based on Farrington’s (2003) framework, Kelley (2007) designed a coding sheet to code the five validity types for each study. This meta-analysis followed Farrington’s (2003) method described above, and adopted Kelley’s (2007) existing coding sheet to evaluate the methodological quality of studies (See Appendix C *Study Methodological Quality Measuring Sheet*). An average score was obtained to represent the overall validity status of each study. Later, analyses for moderators were conducted to investigate the relationship between study quality and effect sizes.

3.7 Effect Size Calculations and Reporting

Effect size is a set of indices that indicate the magnitude of a treatment effect. Different types of effect size indices have been developed for different formats of study outcome measures. In this meta-analysis, Hedges's *g* was chosen as the effect size index for continuous outcome measures; whereas the Cox index (Cox, 1979) was used as the default effect size measure for dichotomous outcomes. The rationales and computation formula are discussed in the following two sections.

3.7.1 Continuous Outcomes

Hedges's *g*, as well as Glass's Δ and Cohen's *d* are three most common effect size indices that can be used to denote standardized mean differences. Standardized mean difference (represented by Hedges's *g*) was chosen to indicate the effects of acceleration in this meta-analysis because there are no universal metrics with which to measure the diverse study variables, such as IQ, achievement scores, or social-emotional development in existing acceleration research. In other words, it is through standardized differences in effect sizes that it is possible to compare effects across different research studies in which different measures have been used. Standardized mean differences can be calculated with a generic formula: $\{M_E - M_C\}/SD$.

This formula can be interpreted that the effect size is obtained through using the mean of the experimental group (M_E) minus the mean of the comparison group (M_C) divided by the standard deviation (SD). There are different options with respect to the usage of SD as the denominator. Some researchers have suggested using the SD of the comparison group. Cohen (1969) proposed the use of a pooled standard deviation (PSD), which is to pool the standard deviation across both the experimental and the comparison group. In this meta-analysis, Cohen's PSD was selected for four reasons: (a) the PSD is more stable and provides a better estimate of the population variance than the comparison group SD (Hedges & Olkin, 1985; Hunter & Schmidt, 1990; Rosenthal, 1991), (b) by using the PSD, effect sizes calculated or estimated from inferential statistics (e.g., *t*-values, *F*-values, ANOVA tables, or *p*-values) are more readily comparable to the directly calculated effect sizes from descriptive statistics, (c) the PSD is a more appropriate measure when it is unclear which group is the comparison condition (Lou, Abrami, Spence, Poulsen, Chambers, & d□Apollonia, 1996) and this may happen in some acceleration studies, and (d) the PSD is based on a larger sample size with the combined experimental groups and comparison groups. This feature is especially beneficial for this meta-analysis of the effects of acceleration, since acceleration research, as well as gifted education research overall, often has small populations and sample sizes (Asher, 1986; Rogers, 1991).

Using the PSD as the denominator, the effect size (Hedges's *g*) computation formula became:

$$g = \{M_E - M_C\}/\{SQRT[(SD_E^2+SD_C^2)/2]\}$$

Specifically, for this meta-analysis, effect sizes were computed using the total of the mean of the experimental group (accelerated group) minus the mean of the comparison group (non-accelerated group) divided by the pooled standard deviation (PSD).

The preference for Hedges's *g* over other alternative indices, such as Cohen's *d* and Glass' *Δ,* was based on the fact that Hedges's *g* can be corrected in order to reduce the bias that may arise when the sample size is very small (i.e., n< 40). This was an important advantage for this study because the samples in many studies of high-ability learners are quite small. A simple correction for the bias from small sample size so that an unbiased effect size estimate can be obtained involves multiplying Hedges's *g* by a factor of [1-3 / (4N-9)] (Hedges, 1981). Formulaically,

$$\text{Corrected } g = g \bullet [1-3/(4N-9)],$$

where N is the total sample size of the study. Unless otherwise noted, Hedges's *g* (particularly corrected Hedges's *g*) is the default effect size measure for the continuous outcomes in this meta-analysis. In other words, Hedges's *g* generally refers to the actual corrected Hedges's *g* from hereon.

3.7.2 Dichotomous Outcome

The Cox index (Cox, 1970) was used as the default effect size measure for dichotomous outcomes. The rationale for this decision began with Odds Ratio (OR). OR is based on the notion of odds, which indicate the occurrence proportion of an event. Odds are calculated through the value of *p*, the probability of the occurrence of an event within a certain group. In the case of this meta-analysis, they should be within treatment group or comparison group. The probabilities of occurrence of an event for the treatment group and the comparison group are denoted as p_1 and p_2 respectively. The corresponding computation formulas for OR are as follows:

$$\text{Odds} = p/(1-p)$$

$$\text{Odds}_1 = p_1/(1-p_1),\ \text{Odds}_2 = p_2/(1-p_2)$$

$$\text{OR} = \text{Odds}_1/\text{Odds}_2$$

Therefore, $\text{OR} = \text{Odds}_1/\text{Odds}_2 = [p_1(1-p_2)]/[p_2(1-p_1)]$

The Odds Ratio (OR) is recommended as the effect size index for studies with dichotomous outcomes, because OR shows statistical and practical advantages over some other alternative effect size indices, such as the ratio of two probabilities, the difference between two probabilities, or the phi coefficient (Fleiss, 1994; Lipsey & Wilson, 2001). In order to simplify statistical analyses, a conventional practice is to transform the OR to Logged Odds Ratio (LOR) (i.e., the natural log of the OR). That is:

$$\text{LOR} = \ln(\text{OR})$$

Furthermore, to make LOR comparable to Hedges's *g* (which was the selected effect size index for continuous outcomes for this meta-analysis), the Cox index (Cox, 1970) was used as the effect size index for dichotomous outcomes. As suggested by Cox (1970), the Cox index (LOR_{Cox}) can be calculated with the following formula:

$$LOR_{Cox}=LOR/1.65$$

The LOR_{Cox} produces effect size values very similar to the values of Hedges's *g*, therefore, the effect size values indicated by LOR_{Cox} and those indexed as Hedges's *g* are comparable. The transformation of LOR to LOR_{Cox} is a highly recommended procedure, for the reason that LOR_{Cox} is found to be the least biased estimator for the standardized mean difference of populations (see Sanchez-Meca, Marin-Martinez, & Chacon-Moscoso, 2003). This procedure has been accepted as the preferred method in research synthesis practice. For instance, the What Works Clearinghouse (WWC) claimed that it “has adopted the Cox index as the default ES measure for dichotomous outcomes” (What Works Clearinghouse, 2007, p.11).

Although both forms of effect sizes (i.e., Hedges's *g* and LOR_{Cox}) showed up during the effect size extraction due to the diverse study outcome formats, Hedges's *g* was chosen as the primary effect size index in this meta-analysis. There were three reasons for doing so: (a) the majority of the effect sizes extracted from the included studies were Hedges's *g*, (b) the values of Hedges's *g* and LOR_{Cox} are comparable as discussed above, and (c) the Comprehensive Meta-Analysis (CMA) software provided various effect size formats (including both Hedges's *g* and LOR_{Cox}) for every effect size, so choosing one of them to present another is feasible. Therefore, as the primary effect size index, Hedges's *g* was chosen as the only effect size format for results reporting in this meta-analysis.

3.8 Effect Size Extraction, Combination and Categorization

In the preliminary stage, effect sizes or sub-effect sizes were extracted for all the constructs or outcome variables reported in the primary studies. For example, only one effect size was extracted from the Lupkowski, Whitmore and Ramsay (1992) study because this study only provided results about a single social-emotional construct (i.e., self-esteem). The Weiner (1985) study only measured the academic performance of accelerated and non-accelerated gifted students. Therefore, a single effect size for academic performance was obtained. However, the Brody (1985) study provided results both for SAT-verbal and SAT-math scores. Therefore, two sub-effect sizes were obtained for this study. With a few exceptions (e.g., Hsu, 2003; Callahan & Smith, 1990; Lupkowski, Whitmore & Ramsay, 1992; Ysseldyke, Tardrew, Betts, Thill & Hannigan, 2004), multiple or at least two sub-effect sizes were extracted for each of the primary studies included. Effect sizes extracted at this stage are called preliminary effect sizes.

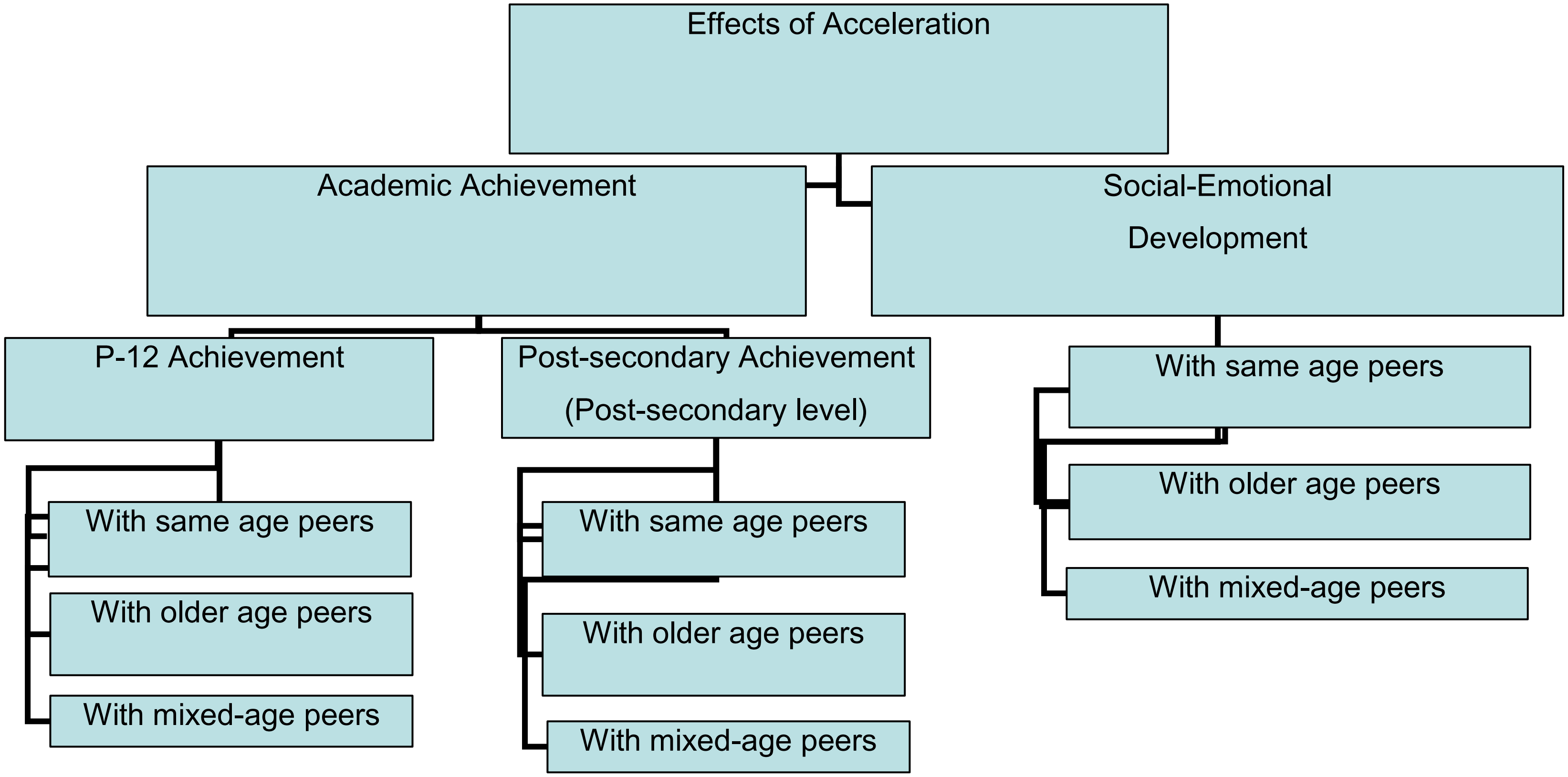
Multiple preliminary effect sizes obtained from a single primary study raises the issue of effect size dependence. This can be problematic because a meta-analysis is normally based on the assumption of independent effect sizes. The Type I error of the homogeneity of effect size tests would increase, if the dependence was ignored and the within group effects treated as if they were independent (Gleser & Olkin, 1994). However, the likelihood of a Type II error would increase if the data were discarded. As recommended by Hedges, Shymansky and Woodworth (1989), the preliminary effect sizes were averaged within each primary study. This procedure is called effect size combination. It was performed after all the preliminary effect sizes were extracted.

The effect sizes were then grouped by each major outcome category (i.e., academic achievement and social-emotional development) of the effects of acceleration on high-ability learners. When a single study provided results both for academic achievement and social-emotional development, two overall effect sizes were yielded. For example, the Ambruster (1995) study produced two effect sizes, one for academic achievement, and another for social-emotional development effects. All the effect sizes for academic achievement consisted of the data for the meta-analysis of the effects of acceleration on academic achievement. Likewise, all the effect sizes related to social-emotional development were the data for a meta-analysis of the social-emotional effects.

For academic achievement, study outcomes were categorized into two developmental levels: P-12 Achievement and Post-secondary Achievement, in order to provide more specific information about academic achievement effects in terms of education levels. P-12 Achievement referred to the typical achievement outcomes at the P-12 level, such as student SAT scores, cumulative Grade Point Average (GPA), number of awards earned, composite score on achievement tests, subject test scores, etc. Post-secondary Achievement denoted achievement outcomes at the post-secondary level, such as university credits earned, college GPA, mean age when graduated from college, graduating with honors, pursuing graduate degrees, etc. The outcomes of two studies (i.e., Barnett & Durden, 1993; Brody, Assouline, & Stanley, 1990) consisted of both P-12 Achievement and Post-secondary Achievement levels.

Moreover, separate effect sizes were calculated in terms of the status of comparison groups in the primary studies. Kulik and Kulik (1984), as well as Kulik (2004), conducted their meta-analyses on the effects of acceleration for two sets of studies: one set was studies with same age control groups and another set was studies with older age control groups. This approach was adopted and used in this meta-analysis. Therefore, the effect sizes were grouped into two subgroups: one with same age peers and another with older age peers. However, as the study coding proceeded, new situations arose: some studies either provided no specific information about the age of the comparison groups, or reported the results with roughly set treatment and comparison groups. For example, some studies included all the subjects from first to 5th grade accelerated gifted students into treatment group, comparing them with the non-accelerated gifted students from first to 5th grade, with no attention paid to the age factors. In this kind of situations, the studies could not fit into the subgroup of with same age peers, nor that of with older age peers. To solve this issue, a third subgroup— mixed-age peers— was created and utilized to categorize studies that compared groups of students with wide age ranges. In summary, outcomes were categorized into three subgroups: same age peers, older age peers, and mixed-age peers. Figure 1 depicts a graphic summary of the study outcomes in this meta-analysis.

Figure 1. Study Outcomes



3.9 Analysis of the Effect Size Data

Data analyses were conducted with the Comprehensive Meta-Analysis (CMA) software (Borenstein, Hedges, Higgins, & Rothstein, 2006), using studies as the unit of analysis, under random effects models. CMA is a specialized meta-analysis software program, which can be used to conduct analyses with both fixed and random effects models. It also scores highest on usability and includes the most complete set of analytical features among existing meta-analysis software (Bax, Yu, Ikeda, & Moons, 2007). Furthermore, it is a user-friendly software program (Pierce, 2007).

Generally, the choice of analysis models is primarily based on the nature of the models and the underlying assumptions of a particular meta-analysis (Borenstein, Hedges, & Rothstein, in press). A fixed effects model functions with the assumption that there is one true effect in all the studies included in a meta-analysis, and the combined effect will be an estimate of that value. In addition, the studies included in a meta-analysis under a fixed effects model should have enough in common in most of the important features. A random effects model, however, assumes the there is more than one true effect. In other words, the combined effect size in a meta-analysis is the mean of a distribution of effect size values. It is believed that, under a random effects model, the studies included in a meta-analysis are drawn from populations that can differ from each other substantially. As a result, these differences could influence the treatment effect.

A random effects model was chosen for this meta-analysis. There were two reasons for doing so. First, it is hard to justify the presumption that there is only one true effect for the effects of acceleration on high ability learners, as a fixed effects model might assume, because the existing literature shows that there may be other significant additional sources of variance in effect sizes resulting from the factors other than the sampling errors. In other words, there are many variables, other than sampling errors, that could significantly affect or even determine the effects of acceleration on high-ability learners. These variables may include, for instance, the types of educational settings in which the studies were conducted, the length of acceleration program, or the sample sizes. Therefore, the true effect of acceleration on high-ability learners could be a distribution of effect values, rather than a single one. Second, the study coding of this meta-analysis revealed that there was a great deal of variation among the studies of the effects of acceleration in terms of study features. It is not appropriate to choose a fixed effects model under such circumstances. In summary, a random effects model fits this meta-analysis much better than a fixed effects model.

According to Cooper (2007), there are unweighted and weighted procedures for computing the combined effect sizes; in the unweighted procedure, equal weight is given to each effect size, whereas in the weighted procedure, the combined effect size is obtained in a way that greater weight is given to the effect sizes from larger samples. Therefore, it is sound to expect that the average effect sizes obtained in this way would be better than those from unweighted procedures because larger samples can provide more precise information about the population. Hence, weighted procedures were used in this study to calculate combined (average) effect sizes. The actual computation operation was performed with the Comprehensive Meta-Analysis (CMA) (Borenstein, Hedges, Higgins, & Rothstein, 2006) software.

Confidence intervals were provided when reporting the combined effect sizes, indicating the dispersion of the effect sizes. Vacha-Haase and Thompson (2004) summarized three reasons for recommending confidence intervals when reporting effect sizes: (a) confidence intervals allow researchers to compare a large set of effects across studies in an economical manner, (b) the widths of the confidence intervals enable researchers to assess the precision of the estimates, and (c) the reference to the intervals across studies can help researchers obtain an accurate estimate of the related parameters.

Data analyses also included heterogeneity analysis and testing for moderators. The purpose of heterogeneity analysis is to examine why effect sizes vary from one study to another (Cooper 2007; Hedges et al., 1989) and to determine whether sampling error alone might be responsible for the variance among the effect sizes. More specifically, heterogeneity analysis is used to investigate whether or not the within-study variance (sampling error) is responsible for the total variance. In other words, heterogeneity analysis determines if between-study variance was significantly associated with the total variance. The total variance of a group of effect sizes that are used to calculate the combined effect size comes from two sources: within-study variance and between-study variance. Within-study variance denotes the variance that is caused by sampling error alone, while between-study variance is caused by the variation across the effect sizes that are associated with different studies. A set of statistics were computed through the CMA software to provide heterogeneity information. These included: the Q-statistic, the degrees of freedom (*df*) of the Q-statistic, *p*-values, and I-squared (I^2) values.

A Q-statistic represents the variance associated with a group or sub-group of effect sizes that are used to calculate the combined effect size for the group or sub-groups. A *p*-value denotes the result of testing heterogeneity for a group or sub-groups of effect sizes. The null hypothesis for a heterogeneity test would be that between-study variance is not significantly associated with the total variance. A significant *p*-value of a heterogeneity test suggests that the null hypothesis can be rejected and it can be concluded that between-study variance is significantly associated with the total variance. In other words, it can be said that there is a high degree of heterogeneity among the group of effect sizes tested.

I-squared values range from 0 to 100. It represents the percentage of the variance that is due to between-study variance. Thus, it quantitatively indicates the degree of heterogeneity. The higher the I-squared value, the higher degree of heterogeneity is present across the effect sizes (studies). A rough guideline proposed by Borenstein, Hedges, and Rothstein (in press) was utilized in this meta-analysis to interpret the I-squared values: an I-squared value of 25 indicates low, a value of 50 suggests moderate and a value of 75 represents a high degree of heterogeneity. I-squared values give additional information for judging the heterogeneity status. If a high degree of heterogeneity exists, further analyses would be conducted to search for possible moderators.

The test for moderators refers to a process of calculating the combined effect sizes for the subsets of studies to identify if there are factors that have influence on the targeted relationship. In this meta-analysis, tests for moderators were conducted to explore the influence of certain variables (moderators) on the combined effect for a group or sub-group of studies. When an ANOVA analog is used to test moderators, the potential moderators are always categorical variables. The statistics relevant to testing for moderators include: the number of studies for each variable and its associated coding categories, the point estimate of effect size, the 95% confidence interval, Q_w, the degrees of freedom (*df*) of Q_w, I-squared values, Q_b and *p*-values. Q_w and I-squared values indicate the degrees of heterogeneity within each sub-group of the effect sizes when the effect sizes (studies) are grouped by the coding categories of a certain variable. Q_b statistic denotes the variance that is caused by the variation between/among all the sub-groups of the effect sizes when these effect sizes are grouped by the coding categories for a certain variable. The *p* -value denotes the results of significance test about the difference between or among the sub-groups.

3.10 Assessment of Publication Bias

Two approaches were used to detect and correct for the issue of publication bias in this meta-analysis. First, a funnel plot was produced and evaluated. In this funnel plot, the studies' effect sizes were plotted against their precision, which was the inverse of the standard error. The diagnostic function of a funnel plot is based on the key idea that if the studies distribute symmetrically around the mean effect size, it can be concluded that there is no publication bias; if more studies (usually small studies) asymmetrically cluster in the bottom of the plot, there might be publication bias.The underlying assumption for the presence of publication bias is that when more small studies asymmetrically concentrate in the bottom of the plot, it indicates the fact that there are more smaller studies published for the reason that they usually produces effects larger than the average, which enables them more likely to meet the criterion for publication.

Second, a trim and fill procedure, was conducted to further assess and adjust for publication bias. The trim and fill procedure was developed by Duvall and Tweedie (2000) and was offered as one of the features in the CMA software. Through this procedure, the CMA removes (trim) unmatched observations from the funnel plot, add (fill) imputed values for likely missing studies, recalculate the combined effect sizes. The appearance of too many missing studies on one side of the mean effect line would suggest that there is presence of publication bias.

CHAPTER 4 DESCRIPTIVE RESULTS

This chapter presents the descriptive results of this meta-analysis. Specifically, the results about the following five topics are reported: (a) study inclusion and exclusion, (b) study coding and outcome variables, (c) general characteristics of the included studies, (d) methodological features of the included studies, and (e) study outcomes and extracted effect sizes.

4.1 Study Inclusion and Exclusion

Following the study search procedures described in Chapter 3, a total of 109 primary studies were located. However, only 38 studies met all the inclusion criteria and were included in this meta-analysis. Table 4 presents the study retrieval sources and the number of studies/percentage from each study source. Of the 38 studies, 25 (65.8%) were published in peer-reviewed journals, including the four major journals in gifted education (i.e., *Gifted Child Quarterly, Roeper Review*, *Journal for the Education of the Gifted*, *Gifted Education International)*; 10 (26.3%) studies were doctoral dissertations; 3 (2.6%) studies were current research, which were obtained from researchers in gifted education through personal contacts.

Table 4. Studies Included in Meta-Analysis

Study Source	No. of Studies	Percentage
Gifted Child Quarterly	11	29%
Roeper Review	4	10.5%
Journal for the Education of the Gifted	2	5.3%
Gifted Education International	1	2.6%
Other journals	7	18.4%
Current research	3	7.9%
Doctoral dissertations	10	26.3%
Total	38	100%

To provide an overview of the time range that the included studies were conducted over, studies were broken into three groups: the group of 1980s, covering the studies that were conducted between the years of 1984 to1989; the group of 1990s, including those conducted between the years of 1990 to 1999; the group of 2000s, consisting of those conducted between the years of 2000 to 2008. Table 5 shows the number of studies and percentage of the studies in each time period.

Table 5. Included Studies and Time Distribution

Time	No. of studies	Percentage
Year 1984---Year 1989	7	18.4%
Year 1990---Year 1999	17	44.8%
Year 2000---2008	14	36.8%
Total	38	100%

The included 38 studies survived multiple screenings. Seventy-one studies (out of the total of 109 studies) were eliminated during the search process because information in abstracts or overviews indicated they did not meet the inclusion criteria. Among the 71 deleted studies, 57 studies were excluded based on their abstracts and 14 studies were excluded for similar reasons after being given a careful examination of their full texts. Table 6 presents the included 38 studies’ important study features, the effect sizes for both social/emotional and academic outcomes. Table 7 presents the 14 studies that were excluded in the final stages, along with the reasons they were excluded.

Table 6. Included Studies and Their Effect Sizes

Study	Comparison group	Study outcome	Sample size	Effect size
Ambruster, 1995	mixed-age peers	Social-emotional Development	200	-0.193
	mixed-age peers	P-12 Achievement	200	-0.022
Barnett & Durden, 1993	mixed-age peers	P-12 Achievement	414	0.306
	mixed-age peers	Post-secondary Achievement	414	0.294
Brody & Benbow, 1987	mixed-age peers	Social-emotional Development	252	-0.023
	mixed-age peers	P-12 Achievement	238	0.315
Brody, 1985	same age peers	P-12 Achievement	126	0.039
Brody, Assouline, & Stanley, 1990	older age peers	P-12 Achievement	3120	0.436
	older age peers	Post-secondary Achievement	2709	0.971
Callahan & Smith, 1990	same age peers	P-12 Achievement	20	1.671
Chilton, 2001	same age peers	Social-emotional Development	141	0.1047
	same age peers	P-12 Achievement	161	0.25

Study	Comparison group	Study outcome	Sample size	Effect size
Cornell, Callahan, & Loyd, 1991	same age peers	Social-emotional Development	51	0.11
Fowler, 2007	same age peers	P-12 Achievement	120	0.206
Gagné & Gagnier, 2004	older age peers	Social-emotional Development	251	0.174
	older age peers	P-12 Achievement	251	0.301
Houston, 1999	same age peers	Social-emotional Development	39	0.0825
Hsu, 2003	mixed-age peers	P-12 Achievement	784	-2.493
Ingersoll, 1992	older age peers	Social-emotional Development	39	-0.171
Janos & Robinson, 1985	older age peers	Social-emotional Development	42	0.574
	older age peers	Post-secondary Achievement	48	0.287
Janos, 1987	same age peers	Post-secondary Achievement	38	1.809

Study	Comparison group	Study outcome	Sample size	Effect size
Janos, Robinson, & Lunneborg, 1989	older age peers	Social-emotional Development	85	-0.124
	older age peers	Post-secondary Achievement	82	-0.35
	same age peers	Social-emotional Development	87	0.144
Jin & Moon, 2006	same age peers	Social-emotional Development	273	0.664
Lupkowski, Whitmore, & Ramsay, 1992	same age peers	Social-emotional Development	218	0.163
Ma, 2002	mixed-age peers	Social-emotional Development	276	0.063
Ma, 2005	mixed-age peers	P-12 Achievement	276	-0.02
Mills, Ablard, & Gustin, 1994	same age peers	P-12 Achievement	332	0.567
Moon, & Callahan, 2001	mixed-age peers	P-12 Achievement	76	0.036

Study	Comparison group	Study outcome	Sample size	Effect size
	older age peers	Social-emotional Development	92	0.046
Noble & Robinson, 1993	older age peers	Post-secondary Achievement	93	0.067
	same age peers	Social-emotional Development	97	0.048
Plucker, & Taylor, 1998	same age peers	Social-emotional Development	601	0.041
Reis, Eckert, McCoach, Jacobs, & Coyne, 2008	same age peers	P-12 Achievement	425	-0.629
Reis, Westberg, Kulikowich, & Purcell, 1998	same age peers	P-12 Achievement	280	0.353
Reynolds, 1993	older age peers	Social-emotional Development	68	0.437
	older age peers	P-12 Achievement	63	0.35
Richards, 2006	same age peers	Post-secondary Achievement	4418	0.099
Robinson & Janos, 1986	older age peers	Social-emotional Development	47	-0.084
	same age peers	Social-emotional Development	51	0.01

Study	Comparison group	Study outcome	Sample size	Effect size
Sayler, 2008	older age peers	Social-emotional Development	113	-0.306
Sayley & Brookshire, 1993	older age peers	Social-emotional Development	708	0.024
	older age peers	P-12 Achievement	699	0.081
Stamps, 2004	same age peers	Social-emotional Development	69	0.168
Swiatek & Benbow, 1991a	same age peers	Social-emotional Development	95	-0.528
	same age peers	Post-secondary Achievement	95	0.343
Swiatek, & Benbow, 1991b	older age peers	Social-emotional Development	214	0.053
	older age peers	P-12 Achievement	214	0.206
Washington, 1999	same age peers	Social-emotional Development	181	0.16
	same age peers	P-12 Achievement	174	0.594
Weiner, 1985	same age peers	Post-secondary Achievement	121	0.324
Wells, Lohman, & Marron, 2008	older age peers	P-12 Achievement	34680	0.652

Study	Comparison group	Study outcome	Sample size	Effect size
Ysseldyke, Tardrew, Betts, Thill, & Hannigan, 2004	same age peers	P-12 Achievement	100	0.442

Note. Included studies were indicated with asterisks in the reference list.

Table 7. Illustrative Sample of Excluded Studies and Exclusion Reasons

Study	Exclusion reasons			
	No appropriate comparison groups	No statistical control	Effect size not-calculable	No cause-effect relationship
Black, 1997	X	X		
Frunzi, 1995	X			
Harrington, 2005	X	X		
McCluskey, Baker & Massey, 1996	X		X	
Menzel, 2006	X		X	
Obrzut, Nelson, & Obrzut, 1984	X			
Parker, 1996	X			
Pischke, 2005	X	X		

Study	Exclusion reasons			
	No appropriate comparison groups	No statistical control	Effect size not-calculable	No cause-effect relationship
Pyryt, 1993			X	
Richardon & Benbow, 1990	X		X	
Sayler, 1996	X			
Schrage, 2007	X	X		
Sethna, Wickstrom, Boothe & Stanley, 2001	X		X	
Weitz, 1985	X			X

4.2 Study Coding and Outcome Variables

Study coding was conducted in two steps. For each study, the information presented was first recorded in text in Microsoft excel spread sheet as raw data. The raw data were then coded into numerical data, following *the Study Feature Coding Descriptors* (see Appendix B). For example, for the Chilton (2001) study, the study focus variable was first recorded as “both academic achievement and social-emotional development” in the raw data sheet. Then, a number of “3” was given to this variable. Later, the raw data were rechecked to examine the coding accuracy after all the coding was finished. No major error was found. The time it took to code one study ranged from one to eight hours. The total time cost for coding all of the 38 studies was 103.7 hours. The average time for coding one study was 2.73 hours. The coding reliability between the two coders was .75.

Study coding revealed major outcome variables that appeared in the primary studies of the effects of acceleration. For academic achievement, the most common outcome variables were standardized achievement test results, college GPA, education background (e.g., degrees obtained, the status of higher-education institutions they attended), ages when certain degrees were obtained or when reaching some career achievement, and career status. For social-emotional development, the most widely studied variables were social relationship, participation in extracurricular activities, mental maturity/adjustment, locus of control, flexibility, satisfaction for life, self-concept, self-acceptance, self-reliance, self-esteem, self-confidence, and educational/vocational plans. A detailed recording of the study outcome variables, and the extracted effect size information are presented in Appendices D & E.

4.3 General Features of the Included Studies

General features refer to the major characteristics of the included primary studies in this meta-analysis. Table 8 provides an overview of the general features. As mentioned in the study inclusion and exclusion section above, the

Table 8. General Features of the Included Studies

Variable	Coding Categories	No. of Studies	Percentage
Study Focus	1. Academic achievement	15	39.5
	2. Social-emotional development	11	28.9
	3. Both	12	31.6
Identification	1. Achievement scores	13	34.2
	2. IQ	0	0
	3. Age	1	2.6
	4. Teacher/Parent identification	1	2.6
	5. Self-reported	4	10.5
	6. Multiple ways	13	34.2
	7. Enrolled in gifted/enrichment program/class	2	5.3
	8. Classified as G/T by state, district, schools, etc	2	5.3
	9. Qualified for certain class, school, college, or grade level	2	5.3
Acceleration Categories	1. Grade-based acceleration	18	47.4
	2. Subject-based acceleration	16	42.1
	3. Both 1 and 2	1	2.6
	4. Non-specific	3	7.9
Acceleration Forms	1. Early kindergarten, school or college entrance	14	36.8
	2. Grade skipping	1	2.6

Variable	Coding Categories	No. of Studies	Percentage
	3. Advanced placement/Dual Credits/ International Baccalaureate	2	5.3
	4. Subject-matter acceleration (e.g., math and reading)	10	26.3
	5. Curriculum compacting	2	5.3
	6. Multiple forms	6	15.8
	7. Early graduation	1	2.6
	8. Mentoring	2	5.3
Acceleration Duration	1. Under 14 weeks	2	5.3
	2. 14 to 28 weeks	5	13.2
	3. Over 28 weeks	26	68.4
	4. Non-specific	5	13.2
Grade Level at Acceleration	1. Elementary	8	21.1
	2. Middle school	2	5.3
	3. High school	8	21.1
	4. Post secondary	10	26.3
	5. Both 1 and 2	3	7.9
	6. Both 2 and 3	3	7.9
	7. Both 3 and 4	0	0

Variable	Coding Categories	No. of Studies	Percentage
	8. Kindergarten	2	5.3
	9. varying (include multiple levels)	2	5.3
Study Time	1. Study conducted within 1 year after acceleration	16	42.1
	2. Longitudinal studies	19	50
	3. Both	3	7.9
School type	1. Rural	1	2.6
	2. Suburban	4	10.5
	3. Urban	5	13.2
	4. Not applicable	24	63.2
	5. National	4	10.5
SES	1. Low	3	7.9
	2. Medium	4	10.5
	3. High	2	5.3
	4. Non-specific	26	68.4
	5. Low to medium	3	7.9
	6. Medium to high	0	0

Variable	Coding Categories	No. of Studies	Percentage
Ethnicity	1. Non-minority students dominant (over 60%)	6	15.8
	2. Minority students dominant (over 60%)	3	7.9
	3. Approximately equivalent	2	5.3
	4. Non-specific	27	71.1
Gender	1. Male predominant (over 60%)	9	23.7
	2. Female predominant (over 60%)	8	21.1
	3. Approximately equivalent	10	26.3
	4. Non-specific	11	28.9

Note. Not all of the variables that are listed in *the Study Feature Coding Descriptors (see* Appendix B) are shown here. Specifically, the variables that were found to have too much missing information during the coding are not presented in this table.

total number of studies included in this meta-analysis was 38. It should be noted, though, that Table 8 does not present all of the general feature variables that were coded according to the coding protocol (see Appendix B) because some variables had too much missing information to be included.

As shown in Table 8, among the 38 studies included, 15 (39.5%) were conducted to investigate the academic achievement effects of acceleration on high-ability learners, 11 (28.9%) were carried out to examine the effects of social-emotional development, whereas 12 (31.6%) provided results for both kind of effects. In terms of identification approaches for high-ability learners, achievement scores and multiple methods (e.g., identification through achievement scores, IQ test scores, age, teacher/parent nomination, observation and interviews) were the most common ones, with 13 (34.2%) each. It is noteworthy that no study was found that used only IQ test scores for identification.

When the 38 studies were categorized into four different groups in terms of acceleration forms, 18 (47.4%) of them investigated grade-based acceleration, 16 (42.1%) examined subject-based acceleration, 1 (2.6%) study provided results for both grade-based and subject-based acceleration, and 3 (7.9%) studies investigated the acceleration intervention that could not be specified as either grade-based or subject-based acceleration.

Eight groups/types of acceleration were investigated in the 38 studies. They were: early kindergarten /school /college entrance (14 studies), subject-matter acceleration (10 studies), multiple forms (6 studies), advanced placement/dual credits/international baccalaureate (2 studies), curriculum compacting (2 studies), mentoring (2 studies), grade skipping (1 study), and early graduation (1 study). The categorization of these acceleration groups/types was based upon three information resources: (a) the descriptions of the 18 forms of acceleration provided by Colangelo et al. in *A Nation Deceived: How Schools Hold Back America's Brightest Students* (i.e., The Templeton National Report on Acceleration) (Colangelo, Assouline, & Gross, 2004), (b) the terms used by the authors of the primary studies, and (c) the actual content of the acceleration interventions that was described in the primary studies.

Acceleration duration is a variable that indicates the length of acceleration interventions implemented. Twenty-six (68.4%) studies investigated acceleration practices that lasted over 28 weeks, 5 (13.2%) studies examined acceleration interventions that lasted between 14 to 28 weeks, 5 (13.2%) studies had non-specific acceleration durations, and 2 (5.3%) studies reported acceleration durations of under 14 weeks.

With regard to the grade level at acceleration, 10 (26.3%) studies investigated acceleration at the post-secondary level; 8 (21.1%) studies examined acceleration at the elementary level, 8 (21.1%) at the high school level, 3 (7.9%) at both the elementary and middle school levels, 3 (7.9%) at both middle school and high school level, 2 (5.3%) at kindergarten level, and 2 (5.3%) studies included multiple levels; No study explored both high school and post secondary level.

Study time refers to the amount of time the study was conducted after the acceleration intervention. Nineteen (50%) studies were longitudinal/retrospective studies, which were defined as being conducted more than one year after the acceleration occurred; 16 (42.1%) studies were conducted within one year after acceleration; 3 (7.9%) studies were found with data collected both longitudinally and within one year after acceleration.

School type categorizes the information about the school environment. Twenty-four (63.2%) studies presented no specific information about the school environment in which the studies were conducted; 5 (13.2%) studies reported that the research was conducted in urban schools; 4 (10.5%) were conducted in suburban schools; in 4 (10.5%) studies national databases were utilized; 1 (2.6%) study was conducted in a rural school environment.

Socioeconomic status (SES) is the variable indicating the average socioeconomic status of the study subjects. Twenty-six (68.4%) studies presented no specific information about SES; 4 (10.5%) studies were conducted on subjects with medium SES; 3 (7.9%) studies on subjects with low SES; and 3 (7.9%) studies on subjects with low to medium SES; 2(5.3%) studies had subjects with high SES.

Ethnicity refers to the ethnic group partition of the treatment sample (i.e., the accelerated high-ability students). Twenty-seven (71.1%) studies were conducted with no specific ethnicity information provided; 6 (15.8%) studies were undertaken with non-minority students dominant in treatment samples; 3 (7.9%) studies were carried out with minority students dominant in treatment samples; 2 (5.3%) studies were conducted with treatment samples in which non-minority and minority students were approximately equivalent in partition size.

Gender is the variable representing the sex ratio of the treatment groups (i.e., the accelerated groups). Eleven (28.9%) studies provided no specific information about gender; 10 (26.3%) studies were conducted with treatment groups of approximately equivalent gender populations; 9 (23.7%) studies were carried out on male predominant treatment groups; 8 (21.1%) studies were carried out with female predominant treatment groups.

4.4 Methodological Features of the Included Studies

Methodological features of studies may have some influence on the observed intervention effects and should not be ignored. In addition, methodological features can be potential sources of moderators, and this is one of the issues that this meta-analysis explored. Therefore, it was important to examine the methodological characteristics of the included studies (see Table 9). As discussed in the general features of the included studies section above, Table 9 does not present all of the methodological feature variables that were described in the coding protocol. Some variables that were found to have too much missing information during the coding were not reported in this table.

As Table 9 reveals, in terms of study retrieval sources, 25 (65.8%) studies were published peer-reviewed journal articles, 10 (26.3%) studies were doctoral dissertations, and 3 (7.9%) studies were unpublished current studies obtained from the authors through personal contacts. One of the ways in which this meta-analysis differs from all the previous meta-analytic studies is in the inclusion of a comparatively higher percentage of doctoral dissertations and some ongoing research studies.

Sample assignment is the variable indicating the methods of assigning subjects to treatment and comparison groups in the primary studies. Thirty-one (81.6%) studies used non-randomized methods, 8 (18.4%) studies used randomized methods, and no study used both randomized and non-randomized methods. This information reflects the fact that most of the research on the effects of acceleration is quasi-experimental in nature.

Comparison group construction is the variable that indicates how the comparison groups were constructed in the primary studies. In twenty-two (57.9%) studies comparison groups were constructed by matching IQ, achievement scores, age, and/or grade-level; in 6 (15.8%) studies the researchers used randomization with an initial pool of eligible subjects; in 4 (10.5%) studies the researchers implemented an ex post facto design, which relied on archival data to construct comparison groups; in 3 (7.9%) of the studies the investigators assigned the rest of the eligible subjects to comparison groups after the treatment group had been assigned; in 2 (5.3%) studies the researchers used randomized clustered sample groups; in 1 (2.6%) study the researcher(s) constructed the comparison group by using self-reported information.

Subject group number is the total number of groups (including both treatment and comparison groups) in the study. Among the 38 studies included, 25 (65.8%) studies had two subject groups and 13 (34.2%) studies had more than two subject groups.

Test design is the variable that indicates the types of test designs that were used to measure the study outcomes. Twenty-eight (73.7%) studies used posttest only, 7 (18.4%) studies used pretest-posttest designs, and 3 (7.9%) studies used multiple times tests.

Research design refers to the types of research that was conducted. There were thirty-one (71.6%) causal-comparative studies, 4 (10.5%) quasi-experimental research studies, and 3 (7.9%) experimental research studies. Experimental research on the effects of acceleration is scarce.

Sample size is the variable used to categorize the sample sizes of the primary studies. Four (10.5%) studies were conducted with a sample size larger than 1000;4 (10.5%) studies with a sample size between 501 to 1000;19 (50%) studies were conducted with a sample size between 101 to 500;11 (28.9%) studies with a sample size of less than 100.

Statistical analysis is the variable that indicates the statistical analysis procedures that the researchers utilized to analyze data and obtain the study outcomes. In 26 (68.4%) studies researchers used student-level t-tests, ANOVA, or MANOVA; in 8 (21.1%) studies researchers utilized student-level ANCOVA, MANCOVA, or logistical regression analysis; in 3 (7.9%) studies researchers implemented a hierarchical linear model (HLM) analysis with cluster-level assignment; in 1 (2.6%) study the researcher(s) only reported descriptive statistics. No studies were conducted with cluster-level analysis. It can be seen that for the majority of the studies the results were obtained using relatively simple statistical analysis procedures. Thus, these results may not able provide strong evidence for the intervention effect.

Data source indicates the types of sources that the researchers relied on to collect data or measure outcomes in the primary studies. In 16 (42.1%) studies researchers relied on surveys/questionnaires; in 8 (21.1%) studies researchers used non-standardized tests; in 5 (13.2%) studies researchers used standardized tests; in 3 (7.9%) studies researchers combined standardized, non-standardized tests, and surveys/questionnaires; in 3 (7.9%) studies researchers utilized national databases; in 1 (2.6%) study researchers relied on cumulative records or archival data.

Table 9. Methodological Features of the Included Studies

Variable	Coding categories	No. of studies	Percentage
Study retrieval source	1.Journal article	25	65.8
	2.Doctoral dissertation	10	26.3
	3.Unpublished study obtained from author	3	7.9
Sample Assignment	1. Randomized	8	18.4
	2. Non-randomized	31	81.6
	3. Mixed	0	0
Comparison group construction	1. Matching IQ, achievement scores, age, or grade level.	22	57.9
	2. Ex post facto design	4	10.5
	3. Randomization with initial pool of eligible subjects	6	15.8
	4. Randomization with clustered sample groups	2	5.3
	5. Self-reported information	1	2.6
	6. The rest of eligible subjects after treated groups have been assigned.	3	7.9
Subject group numbers	1. Two groups	25	65.8
	2. More than two groups	13	34.2

Variable	Coding categories	No. of studies	Percentage
Test design	1.Posttest only	28	73.7
	2. Pretest-posttest	7	18.4
	3. Multiple times tests	3	7.9
Research design	1. Experimental	3	7.9
	2. Quasi-experimental	4	10.5
	3. Causal-comparative research	31	71.6
Sample Size	1. ≤ 100	11	28.9
	2. 101-500	19	50
	3. 501-1000	4	10.5
	4. >1000	4	10.5
	5. Not applicable	0	0
Statistical Analysis	1. Descriptive statistics (including dichotomous outcomes)	1	2.6
	2. Student-level T-test, ANOVA, or MANOVA	26	68.4
	3. Student-level ANCOVA, MANCOVA, Logistic Regression Analysis	8	21.1
	4. Cluster-level analysis	0	0

Variable	Coding categories	No. of studies	Percentage
	5. Hierarchical Linear Model (HLM) analysis with cluster-level assignment	3	7.9
Data source	1. Standardized tests	5	13.2
	2. Non-standardized tests	8	21.1
	3. Survey/questionnaires	16	42.1
	4. Combining 1, or 2, or 3, or 5.	3	7.9
	5. Cumulative records or archival data	1	2.6
	6. National databases	3	7.9
Reliability information	1. Yes	31	81.6
	2. No	7	18.4
Measurement reliability status	1. High reliability	19	50
	2. Moderate reliability	13	34.2
	3. Low reliability	1	2.6
	4. Not-available	5	13.2
Study validity status	1. Low validity	2	5.3
	2. Moderate validity	19	50
	3. High validity	17	44.7

Variable	Coding categories	No. of studies	Percentage
ES reporting	1. Yes	15	39.5
	2. No	23	60.5
Effect Size Extraction	1. Calculated from descriptive statistics	29	76.3
	2. Calculated from inferential statistics	7	18.4
	3. Estimated from probabilities	1	2.6
	4. Computed effect sizes in original study	1	2.6

Note. Not all of the variables that are listed in *the Study Feature Coding Descriptors (see* Appendix B) are shown here. Specifically, the variables that were found to have too much missing information during the coding are not presented in this table.

Thirty-one (81.6%) studies reported reliability information; however, 7 (18.4%) studies did not do so. In terms of measurement reliability status, 19 (50%) of the studies were considered to have high measurement reliability because the researchers either used instruments with high reliability, or standardized tests to measure study outcomes; 13 (34.2%) studies were considered to have moderate reliability because researchers developed and used instruments or scales with moderate measurement reliability in these studies; 5 (13.2%) studies were found with no specific information available about the measurement reliability; 1 (2.6%) study was considered to have low reliability because the researcher (s) used low reliability measurement instruments. Although it should be noted that the evaluation of measurement reliability status was not free from subjectivity, the information obtained can still be considered to be informative.

Study validity status denotes five kinds of validity, as proposed by Kelley (2007). They are: internal validity, construct validity, statistical validity, external validity, and descriptive validity. Study validity status was measured using *the Study Validity Measuring Sheet* (see Appendix C). An average score of the five kinds of validity was calculated for each study to indicate its study validity status. A value between 0 and 2 represents low validity, a value between 2.1 and 3 represents moderate validity, and a value between 3.1 and 4 represents high validity. Among the 38 studies included, 2 (5.3%) had low validity;19 (50%) studies had moderate validity;17(44.7%) had high validity. It is relevant to note that this information was based upon the comparisons among all the included studies. Therefore, it only denotes a relative sense of 'high validity', 'moderate validity', or 'low validity' in the pool of included studies

Effect size reporting indicates whether the researchers reported effect sizes in the primary studies. Twenty-three (60.5%) studies had no effect sizes reported, whereas 15 (39.5%) studies had effect sizes available. It can be concluded that reporting effect size was still not common practice in acceleration research.

Effect size extraction is the variable representing the ways that effect sizes were extracted from the primary studies. For 29 (76.3%) studies, effect sizes were calculated from descriptive statistics; for 7 (18.4%) studies, effect sizes were calculated from inferential statistics; for 1 (2.6%) study, effect sizes were estimated from probability information; for 1 (2.6%) study, effect sizes were provided by the primary study and no other information was available for calculation.

4.5 Outcomes Variables and Effect Size Information

As described in the methods section, the CMA software can transform all forms of effect sizes into the selected primary effect size index (i.e., Hedges's *g*) so that the effects can be synthesizable. A total of 274 preliminary effect sizes (i.e., effect sizes before combination) were extracted from the 38 studies. On average, there were 7.2 effect sizes extracted from each study. For academic achievement results (141 effect sizes), the effect sizes ranged from -4.145 to 3.843, including 108 positive effect sizes, 4 zero effect sizes and 29 negative effect sizes. For social-emotional development results (133 effect sizes), the effect sizes ranged from -.746 to 1.281, including 81 positive and 52 negative values. Table 10 presents the study effect sizes distribution. In addition, Appendices D & E provide detailed information about all the study outcome variables and the corresponding effect sizes. The combined effect size for each included study will be reported in Chapter 5.

For academic achievement results, 141 preliminary effect sizes were extracted. The majority of effect sizes (n=79) were calculated with effect size format of independent groups, which relies on the means and standard deviations of both treatment and comparison groups. Forty-one effect sizes were computed with the effect size format of cohort two by two (events), which is based on the number of treatment and comparison events, as well as the total number of participants for treatment and comparison group, respectively. This format is usually used to calculate effect sizes for dichotomous outcomes. The remaining 21 effect sizes were obtained with other effect size formats that are provided in the CMA software. The usage of these formats usually relies on the information of means and standard deviations of paired groups, t statistics, *p*-values, and/or sample sizes.

For social-emotional development results, 133 preliminary effect sizes were calculated. Likewise, the majority of effect sizes (n=119) were calculated with effect size format of independent groups; 8 effect sizes were computed with the effect size format of cohort two by two (events); 6 effect sizes were obtained with other effect size formats.

Table 10. Study Effect Sizes Distribution

Results	No. of studies	ESs Range	ES distribution
Academic achievement	28[a]	(-2.493, 1.809)	5 (17.8%) negative 23 (82.2%) positive .294 is the median
Social-emotional development	22[b]	(-.528, .664)	7 (31.8%)negative 15 (68.2%)positive .058 is the median
Total	38[c]		

[a] The number of studies that provided academic achievement results.
[b] The number of studies that reported social-emotional development results.
[c] The total number of studies included in this meta-analysis.

CHAPTER 5 META-ANALYSIS RESULTS

This chapter reports the results of the two meta-analyses in this study. Specifically, this chapter consists of two parts: the first part reports the results of a meta-analysis on the academic achievement effects of acceleration and the second part reports the results of a meta-analysis on the social-emotional effects. For each major category of outcome, the results are divided into the following five categories: combined effects, heterogeneity analysis, testing for moderators, cumulative analysis, and assessment of publication bias. The interpretation and discussion of these results can be found in Chapter 6.

5.1 Academic Achievement Effects Results

5.1. 1 Combined Effects

A separate meta-analysis was conducted on the effects of acceleration on academic achievement. As discussed in the methods section, data analyses were performed using the CMA software, under random effects models, using studies as the unit of analysis. Weighted effect sizes were produced for the combined (average) effects.

As previously described in the methods section, academic achievement results were broken down into two groups by outcome levels: P-12 Achievement and Post-secondary Achievement, denoting the achievement results for P-12 level and post-secondary level, respectively. Furthermore, for each outcome level (i.e., P-12 Achievement or Post-secondary Achievement) the results were calculated by comparison groups (i.e., with older age peers, with same age peers, and with mixed-age peers). Therefore, the analysis results for academic achievement were presented not only in terms of the overall combined effects, but of the effects for the related sub-groups.

Figures 2 through 6 show the effect size statistics and forest plots generated during the analysis of achievement effects. Statistical information in each figure includes: Hedges's *g*, standard error (SE), lower and upper limits of the 95% confidence interval for each effect size, a forest plot, and the relative weight associated with each study. A forest plot graphically indicates the magnitudes and the dispersion of the effect sizes. In the forest plots shown in Figures 2 through 6, the scales for Hedges's *g* were set at (-2, 2), with zero at the middle line, called the zero line; an effect size falling on the left side of the zero line indicates that the result of a study or group favors comparison condition, whereas an effect size lying on the right side of the zero line illustrates that the result of a study or group favors treatment condition. The horizontal lines in the forest plots denote the confidence intervals (95% in this meta-analysis) of the effect sizes. The bullet in the middle point of the 95% confidence interval lines indicates the combined effect size for each study. The size of the bullets is proportional to the relative weight of the study. A wider confidence interval line suggests that the study has a smaller sample size and lower precision, whereas a narrower confidence interval line is a characteristic of a study that has a larger sample size and/or higher precision. If the confidence interval line crosses the zero line, a significance test of the results will not be statistically significant; otherwise, it will be statistically significant.

The summary information for a sub-group appears under all the studies that belong to the sub-group. The summary information for an overall group is listed under all the studies in the overall group. An open diamond in the forest plot denotes the combined effects for a sub-group, and a closed diamond represents the combined effects of for an overall group. Table 11 summarizes the major statistical information for the combined effects which are presented in Figures 2 through 6.

Figure 2 exhibits the overall combined effect sizes for academic achievement (also see Table 11). Twenty-eight studies that investigated academic achievement effects were included in the analysis. The effect size index, Hedges's *g,* was sorted from lowest to highest. As seen in Figure 2, the effect sizes for the included 28 studies ranged from -2.493 to 1.809, a wide range. The overall combined (average) effect size for academic achievement was slightly positive (*g* .180, 95% CI -.072 to .431, 28 studies). The combined effect was not statistically significant (df = 27, p > .05). The information above suggests there was no statistically significant difference between the treatment and comparison groups in terms of the effects of acceleration on high-ability learners' academic achievement, although the combined effect size of all the included studies in this meta-analysis appeared to be larger than zero (i.e., *g* =.180).

It is relevant to note that among the 28 studies that provided results for academic achievement, three studies yielded obviously bigger (either positively or negatively) effect sizes than did the rest of 25 studies. Specifically, the Hsu (2003) study yielded the lowest effect size (*g* = -2.493); whereas the Janos (1987) study yielded the highest effect size (*g* = 1.809), followed by the study of Callahan and Smith (1990) (*g* = 1.671). A close examination of these three studies revealed that in each of these three studies, only a very limited number of outcome variables was investigated. Specifically, for both the Hsu (2003) and the Callahan and Smith (1990) study, only one outcome variable was measured; for the Janos (1987) study, three outcome variables were measured. The limited number of outcome variables led to a limited number of preliminary effect sizes being extracted. It is possible that when there is a limited number of preliminary effect sizes involved in the combination of the effect size for a study, extremely big (either positively or negatively) values were yielded because of lack of balance between or among the preliminary effect sizes. In addition, it was found that for the Hsu (2003) study, a potential positive preliminary effect size was not calculable because of a lack of information. As a result, only a single negative preliminary effect size (*g* = -2.493) was extracted, and this was the only effect size that could be used to represent this study. This could be another reason that the Hsu (2003) study appeared to have the lowest effect size (*g* = -2.493) among all the included studies. In addition, the Hsu (2003) study had low validity. When the Hsu (2003) study was deleted from the analysis, the combined effect size was increased (*g* .240, 95% CI .128 to .358, 27 studies). In addition, the combined effect was statistically significant (*df* = 26, *p* < .001).

Because in this meta-analysis random effects models were utilized, which assume that the effects may vary across studies and different contexts, the three studies mentioned above were still included in the analysis. It should be noted, though, that these studies had a strong influence on the results and may have yielded anomalous result in certain circumstances. Such circumstances may include when only one of these three studies, especially the Hsu (2003) study, was involved in the analysis of a certain group or sub-groups. When such a case occurs, clarification and explanation will be provided.

Figure 3 displays the combined (average) effect sizes for academic achievement grouped by outcome levels (i.e., P-12 Achievement and Post-secondary Achievement) (also see Table 11). For P-12 Achievement the combined effect size was .147 (95% CI -.174 to .467, 21 studies). For Post-secondary Achievement, the combined effect size was .313 (95% CI -.262 to .889, 7 studies). The significance tests of both the combined effects were not statistically significant at the p< .05 level. In addition, no statistically significant difference in the combined effect sizes was detected between P-12 Achievement and Post-secondary Achievement (*df* =1, p >.05).

Figure 4 presents the combined effect sizes for academic achievement by comparison groups (also see Table 11). The combined (average) effect size for the sub-group of Achievement with Same Age Peers was the highest (*g* .396, 95% CI .029 to .762, 13 studies), followed by the sub-group of Achievement with Older Age Peers (*g* .224, 95% CI -.212 to .660, 9 studies). However, for the sub-group of Achievement with Mixed-Age Peers, the combined (average) effect size was negative (*g* -.323, 95% CI -.842 to .197, 6 studies). None of the combined effects of the sub-groups of Achievement with Older Age Peers and Achievement with Mixed-Age Peers was found to be statistically significant (p > .05). However, for the sub-group of Achievement with Same Age Peers, the effect was found to be statistically significant ($p < .05$). Moreover, no statistically significant difference in the combined effect sizes was detected among the three sub-groups (*df* =2, $p > .05$).

Figure 5 shows the results for P-12 Achievement by comparison groups (also see Table 11). Again, three sub-groups were created. They were: P-12 Achievement with Same Age Peers. P-12 Achievement with Older Age Peers, and P-12 Achievement with Mixed-Age Peers. For the sub-group of P-12 Achievement with Same Age Peers, the combined (average) effect size was .347 (95% CI -.165 to .858, 9 studies). For the sub-group of P-12 Achievement with Older Age Peers, the combined (average) effect size was .334 (95% CI -.290 to .959, 6 studies). For the sub-group of P-12 Achievement with Mixed-Age Peers, the combined (average) effect size was -.320 (95% CI -.936 to .295, 6 studies). None of the combined effects of the sub-groups were found to be statistically significant at the significance level of $p < .05$. Again, no statistically significant difference in the combined effect sizes was detected among the above three sub-groups.

Figure 6 and Table 11 present the results for Post-secondary Achievement when it was grouped by comparison groups. For the sub-group of Post-secondary Achievement with Same Age Peers, the combined (average) effect size was .498 (95% CI -.045 to 1.042, 4 studies). For the sub-group of Post-secondary Achievement with Older Age Peers, the combined (average) effect size was .255 (95% CI -.266 to .776, 4 studies). For the sub-group of Post-secondary Achievement with Mixed-Age Peers, the combined (average) effect size was .294 (95% CI -.667 to 1.255,1 study). None of the combined effects of the sub-group was found to be statistically significant at the significance level of $p < .05$. No statistically significant difference in the combined effect sizes was detected among the above three sub-groups.

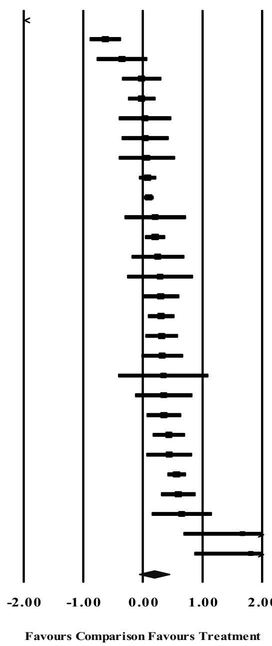
In summary, as displayed in Figures 2 through 6 and Table 11, the overall combined effect size for academic achievement was .180. When Academic Achievement, P-12 Achievement and Post-secondary Achievement were grouped in terms of comparison groups, the sub-groups of with same age peers consistently exhibited higher effect sizes than that of with older and mixed-age peers (i.e., g =.396, g =.347, and g =.498, respectively). This suggests that when accelerated high-ability learners were compared with their same age peers, the academic effects of acceleration were more discernable. Further, the combined effect of the sub-group of Achievement with Same Age Peers was found to be statistically significant ($p < .05$), while the combined effects of all the other group or sub-groups were not statistically significant ($p > .05$).

Figure 2. Overall Combined Effect Size for Academic Achievement

Overall Combined Effect Size for Academic Achievement

Study name	Outcome	Comparison	Hedges's g	Standard error	Lower limit	Upper limit	p-Value
Hsu, 2003	P-12 Achievement	mixed-age peers	-2.493	0.114	-2.716	-2.270	0.000
Reis, Eckert, McCoach, Jacobs, & Coyne, 2008	P-12 Achievement	same age peers	-0.629	0.138	-0.899	-0.359	0.000
Janos, Robinson, & Lunneborg, 1989	Post-secondary Achievement	older age peers	-0.350	0.221	-0.784	0.084	0.114
Ambruster, 1995	P-12 Achievement	mixed-age peers	-0.022	0.173	-0.361	0.317	0.899
Ma, 2005	P-12 Achievement	mixed-age peers	-0.020	0.122	-0.260	0.220	0.870
Moon, & Callahan, 2001	P-12 Achievement	mixed-age peers	0.036	0.230	-0.415	0.487	0.876
Brody, 1985	P-12 Achievement	same age peers	0.039	0.205	-0.363	0.441	0.849
Noble & Robinson, 1993	Post-secondary Achievement	older age peers	0.067	0.245	-0.413	0.547	0.784
Sayley & Brookshire, 1993	P-12 Achievement	older age peers	0.081	0.077	-0.071	0.233	0.296
Richards, 2006	Post-secondary Achievement	same age peers	0.099	0.045	0.011	0.187	0.027
Swiatek, & Benbow, 1991 (2)	P-12 Achievement	older age peers	0.206	0.266	-0.316	0.728	0.439
Fowler, 2007	P-12 Achievement	same age peers	0.206	0.089	0.031	0.381	0.021
Chilton, 2001	P-12 Achievement	same age peers	0.250	0.230	-0.201	0.701	0.278
Janos & Robinson, 1985	Post-secondary Achievement	older age peers	0.287	0.286	-0.274	0.848	0.316
Gagné & Gagnier, 2004	P-12 Achievement	older age peers	0.301	0.161	-0.015	0.617	0.062
Barnett & Durden, 1993	P-12 Achievement	mixed-age peers	0.306	0.118	0.074	0.538	0.010
Brody & Benbow, 1987	P-12 Achievement	mixed-age peers	0.315	0.145	0.031	0.599	0.030
Weiner, 1985	Post-secondary Achievement	same age peers	0.324	0.182	-0.032	0.680	0.074
Swiatek & Benbow, 1991	Post-secondary Achievement	same age peers	0.343	0.390	-0.421	1.107	0.379
Reynolds, 1993	P-12 Achievement	older age peers	0.350	0.249	-0.138	0.838	0.160
Reis, Westberg, Kulikowich, & Purcell, 1998	P-12 Achievement	same age peers	0.353	0.152	0.056	0.650	0.020
Brody, Assouline, & Stanley, 1990	P-12 Achievement	older age peers	0.436	0.141	0.159	0.713	0.002
Ysseldyke, Tardrew, Betts, Thill, & Hannigan, 2004	P-12 Achievement	same age peers	0.442	0.200	0.050	0.834	0.027
Mills, Ablard, & Gustin, 1994	P-12 Achievement	same age peers	0.567	0.084	0.403	0.731	0.000
Washington, 1999	P-12 Achievement	same age peers	0.594	0.152	0.297	0.891	0.000
Wells, Lohman, & Marron, 2008	P-12 Achievement	older age peers	0.652	0.261	0.141	1.163	0.012
Callahan & Smith, 1990	P-12 Achievement	same age peers	1.671	0.511	0.670	2.672	0.001
Janos, 1987	Post-secondary Achievement	same age peers	1.809	0.487	0.855	2.763	0.000
			0.180	0.128	-0.072	0.431	0.162

Hedges's g and 95% CI



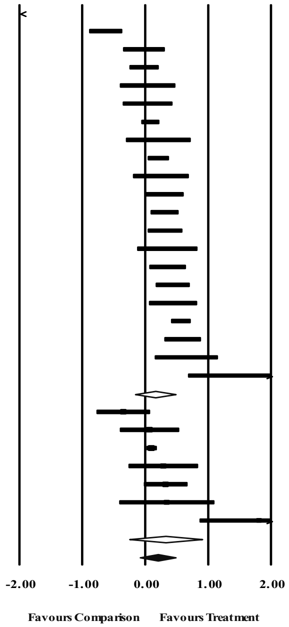
under randomeffects model

Figure 3. Academic Achievement Grouped by Outcome Levels

Academic Achievement Grouped by Outcome Levels

Group by Outcome	Study name	Outcome	Comparison	Hedges's g	Standard error	Lower limit	Upper limit	p-Value
P-12 Achievement	Hsu, 2003	P-12 Achievement	mixed-age peers	-2.493	0.114	-2.716	-2.270	0.000
P-12 Achievement	Reis, Eckert, McCoach, Jacobs, & Coyne, 2008	P-12 Achievement	same age peers	-0.629	0.138	-0.899	-0.359	0.000
P-12 Achievement	Ambruster, 1995	P-12 Achievement	mixed-age peers	-0.022	0.173	-0.361	0.317	0.899
P-12 Achievement	Ma, 2005	P-12 Achievement	mixed-age peers	-0.020	0.122	-0.260	0.220	0.870
P-12 Achievement	Moon, & Callahan, 2001	P-12 Achievement	mixed-age peers	0.036	0.230	-0.415	0.487	0.876
P-12 Achievement	Brody, 1985	P-12 Achievement	same age peers	0.039	0.205	-0.363	0.441	0.849
P-12 Achievement	Sayley & Brookshire, 1993	P-12 Achievement	older age peers	0.081	0.077	-0.071	0.233	0.296
P-12 Achievement	Swiatek, & Benbow, 1991 (2)	P-12 Achievement	older age peers	0.206	0.266	-0.316	0.728	0.439
P-12 Achievement	Fowler, 2007	P-12 Achievement	same age peers	0.206	0.089	0.031	0.381	0.021
P-12 Achievement	Chilton, 2001	P-12 Achievement	same age peers	0.250	0.230	-0.201	0.701	0.278
P-12 Achievement	Gagné & Gagnier, 2004	P-12 Achievement	older age peers	0.301	0.161	-0.015	0.617	0.062
P-12 Achievement	Barnett & Durden, 1993	P-12 Achievement	mixed-age peers	0.306	0.118	0.074	0.538	0.010
P-12 Achievement	Brody & Benbow, 1987	P-12 Achievement	mixed-age peers	0.315	0.145	0.031	0.599	0.030
P-12 Achievement	Reynolds, 1993	P-12 Achievement	older age peers	0.350	0.249	-0.138	0.838	0.160
P-12 Achievement	Reis, Westberg, Kulikowich, & Purcell, 1998	P-12 Achievement	same age peers	0.353	0.152	0.056	0.650	0.020
P-12 Achievement	Brody, Assouline, & Stanley, 1990	P-12 Achievement	older age peers	0.436	0.141	0.159	0.713	0.002
P-12 Achievement	Ysseldyke, Tardrew, Betts, Thill, & Hannigan, 2004	P-12 Achievement	same age peers	0.442	0.200	0.050	0.834	0.027
P-12 Achievement	Mills, Ablard, & Gustin, 1994	P-12 Achievement	same age peers	0.567	0.084	0.403	0.731	0.000
P-12 Achievement	Washington, 1999	P-12 Achievement	same age peers	0.594	0.152	0.297	0.891	0.000
P-12 Achievement	Wells, Lohman, & Marron, 2008	P-12 Achievement	older age peers	0.652	0.261	0.141	1.163	0.012
P-12 Achievement	Callahan & Smith, 1990	P-12 Achievement	same age peers	1.671	0.511	0.670	2.672	0.001
P-12 Achievement				0.147	0.164	-0.174	0.467	0.370
Post-secondary Achievement	Janos, Robinson, & Lunneborg, 1989	Post-secondary Achievement	older age peers	-0.350	0.221	-0.784	0.084	0.114
Post-secondary Achievement	Noble & Robinson, 1993	Post-secondary Achievement	older age peers	0.067	0.245	-0.413	0.547	0.784
Post-secondary Achievement	Richards, 2006	Post-secondary Achievement	same age peers	0.099	0.045	0.011	0.187	0.027
Post-secondary Achievement	Janos & Robinson, 1985	Post-secondary Achievement	older age peers	0.287	0.286	-0.274	0.848	0.316
Post-secondary Achievement	Weiner, 1985	Post-secondary Achievement	same age peers	0.324	0.182	-0.032	0.680	0.074
Post-secondary Achievement	Swiatek & Benbow, 1991	Post-secondary Achievement	same age peers	0.343	0.390	-0.421	1.107	0.379
Post-secondary Achievement	Janos, 1987	Post-secondary Achievement	same age peers	1.809	0.487	0.855	2.763	0.000
Post-secondary Achievement				0.313	0.294	-0.262	0.889	0.286
Overall				0.186	0.143	-0.094	0.466	0.193

Hedges's g and 95% CI



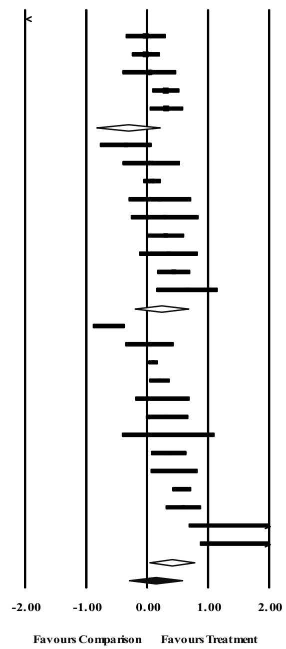
under random effects model

Figure 4. Academic Achievement Grouped by Comparison Groups

Academic Achievement Grouped by Comparison Groups

Group by Comparison	Study name	Outcome	Comparison	Hedges's g	Standard error	Lower limit	Upper limit	p-Value
mixed-age peers	Hsu, 2003	P-12 Achievement	mixed-age peers	-2.493	0.114	-2.716	-2.270	0.000
mixed-age peers	Ambruster, 1995	P-12 Achievement	mixed-age peers	-0.022	0.173	-0.361	0.317	0.899
mixed-age peers	Ma, 2005	P-12 Achievement	mixed-age peers	-0.020	0.122	-0.260	0.220	0.870
mixed-age peers	Moon, & Callahan, 2001	P-12 Achievement	mixed-age peers	0.036	0.230	-0.415	0.487	0.876
mixed-age peers	Barnett & Durden, 1993	P-12 Achievement	mixed-age peers	0.306	0.118	0.074	0.538	0.010
mixed-age peers	Brody & Benbow, 1987	P-12 Achievement	mixed-age peers	0.315	0.145	0.031	0.599	0.030
mixed-age peers				-0.323	0.265	-0.842	0.197	0.223
older age peers	Janos, Robinson, & Lunneborg, 1989	Post-secondary Achievement	older age peers	-0.350	0.221	-0.784	0.084	0.114
older age peers	Noble & Robinson, 1993	Post-secondary Achievement	older age peers	0.067	0.245	-0.413	0.547	0.784
older age peers	Sayley & Brookshire, 1993	P-12 Achievement	older age peers	0.081	0.077	-0.071	0.233	0.296
older age peers	Swiatek, & Benbow, 1991 (2)	P-12 Achievement	older age peers	0.206	0.266	-0.316	0.728	0.439
older age peers	Janos & Robinson, 1985	Post-secondary Achievement	older age peers	0.287	0.286	-0.274	0.848	0.316
older age peers	Gagné & Gagnier, 2004	P-12 Achievement	older age peers	0.301	0.161	-0.015	0.617	0.062
older age peers	Reynolds, 1993	P-12 Achievement	older age peers	0.350	0.249	-0.138	0.838	0.160
older age peers	Brody, Assouline, & Stanley, 1990	P-12 Achievement	older age peers	0.436	0.141	0.159	0.713	0.002
older age peers	Wells, Lohman, & Marron, 2008	P-12 Achievement	older age peers	0.652	0.261	0.141	1.163	0.012
older age peers				0.224	0.222	-0.212	0.660	0.315
same age peers	Reis, Eckert, McCoach, Jacobs, & Coyne, 2008	P-12 Achievement	same age peers	-0.629	0.138	-0.899	-0.359	0.000
same age peers	Brody, 1985	P-12 Achievement	same age peers	0.039	0.205	-0.363	0.441	0.849
same age peers	Richards, 2006	Post-secondary Achievement	same age peers	0.099	0.045	0.011	0.187	0.027
same age peers	Fowler, 2007	P-12 Achievement	same age peers	0.206	0.089	0.031	0.381	0.021
same age peers	Chilton, 2001	P-12 Achievement	same age peers	0.250	0.230	-0.201	0.701	0.278
same age peers	Weiner, 1985	Post-secondary Achievement	same age peers	0.324	0.182	-0.032	0.680	0.074
same age peers	Swiatek & Benbow, 1991	Post-secondary Achievement	same age peers	0.343	0.390	-0.421	1.107	0.379
same age peers	Reis, Westberg, Kulikowich, & Purcell, 1998	P-12 Achievement	same age peers	0.353	0.152	0.056	0.650	0.020
same age peers	Ysseldyke, Tardrew, Betts, Thill, & Hannigan, 2004	P-12 Achievement	same age peers	0.442	0.200	0.050	0.834	0.027
same age peers	Mills, Ablard, & Gustin, 1994	P-12 Achievement	same age peers	0.567	0.084	0.403	0.731	0.000
same age peers	Washington, 1999	P-12 Achievement	same age peers	0.594	0.152	0.297	0.891	0.000
same age peers	Callahan & Smith, 1990	P-12 Achievement	same age peers	1.671	0.511	0.670	2.672	0.001
same age peers	Janos, 1987	Post-secondary Achievement	same age peers	1.809	0.487	0.855	2.763	0.000
same age peers				0.396	0.187	0.029	0.762	0.034
Overall				0.127	0.222	-0.309	0.563	0.568

Hedges's g and 95% CI



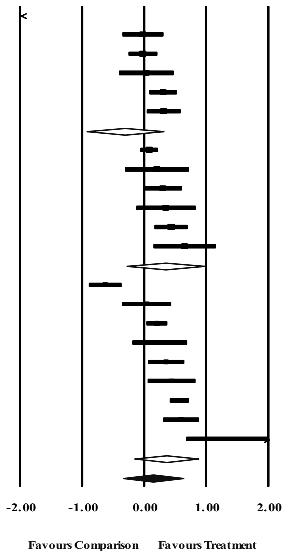
under random effects model

Figure 5. P-12 Achievement Grouped by Comparison Groups

P-12 Achievement Grouped by Comparison Groups

Group by Comparison	Study name	Outcome	Comparison	Hedges's g	Standard error	Lower limit	Upper limit	p-Value
mixed-age peers	Hsu, 2003	P-12 Achievement	mixed-age peers	-2.493	0.114	-2.716	-2.270	0.000
mixed-age peers	Ambruster, 1995	P-12 Achievement	mixed-age peers	-0.022	0.173	-0.361	0.317	0.899
mixed-age peers	Ma, 2005	P-12 Achievement	mixed-age peers	-0.020	0.122	-0.260	0.220	0.870
mixed-age peers	Moon, & Callahan, 2001	P-12 Achievement	mixed-age peers	0.036	0.230	-0.415	0.487	0.876
mixed-age peers	Barnett & Durden, 1993	P-12 Achievement	mixed-age peers	0.306	0.118	0.074	0.538	0.010
mixed-age peers	Brody & Benbow, 1987	P-12 Achievement	mixed-age peers	0.315	0.145	0.031	0.599	0.030
mixed-age peers				-0.320	0.314	-0.936	0.295	0.308
older age peers	Sayley & Brookshire, 1993	P-12 Achievement	older age peers	0.081	0.077	-0.071	0.233	0.296
older age peers	Swiatek, & Benbow, 1991 (2)	P-12 Achievement	older age peers	0.206	0.266	-0.316	0.728	0.439
older age peers	Gagné & Gagnier, 2004	P-12 Achievement	older age peers	0.301	0.161	-0.015	0.617	0.062
older age peers	Reynolds, 1993	P-12 Achievement	older age peers	0.350	0.249	-0.138	0.838	0.160
older age peers	Brody, Assouline, & Stanley, 1990	P-12 Achievement	older age peers	0.436	0.141	0.159	0.713	0.002
older age peers	Wells, Lohman, & Marron, 2008	P-12 Achievement	older age peers	0.652	0.261	0.141	1.163	0.012
older age peers				0.334	0.319	-0.290	0.959	0.294
same age peers	Reis, Eckert, McCoach, Jacobs, & Coyne, 2008	P-12 Achievement	same age peers	-0.629	0.138	-0.899	-0.359	0.000
same age peers	Brody, 1985	P-12 Achievement	same age peers	0.039	0.205	-0.363	0.441	0.849
same age peers	Fowler, 2007	P-12 Achievement	same age peers	0.206	0.089	0.031	0.381	0.021
same age peers	Chilton, 2001	P-12 Achievement	same age peers	0.250	0.230	-0.201	0.701	0.278
same age peers	Reis, Westberg, Kulikowich, & Purcell, 1998	P-12 Achievement	same age peers	0.353	0.152	0.056	0.650	0.020
same age peers	Ysseldyke, Tardrew, Betts, Thill, & Hannigan, 2004	P-12 Achievement	same age peers	0.442	0.200	0.050	0.834	0.027
same age peers	Mills, Ablard, & Gustin, 1994	P-12 Achievement	same age peers	0.567	0.084	0.403	0.731	0.000
same age peers	Washington, 1999	P-12 Achievement	same age peers	0.594	0.152	0.297	0.891	0.000
same age peers	Callahan & Smith, 1990	P-12 Achievement	same age peers	1.671	0.511	0.670	2.672	0.001
same age peers				0.347	0.261	-0.165	0.858	0.184
Overall				0.133	0.247	-0.351	0.616	0.591

Hedges's g and 95% CI



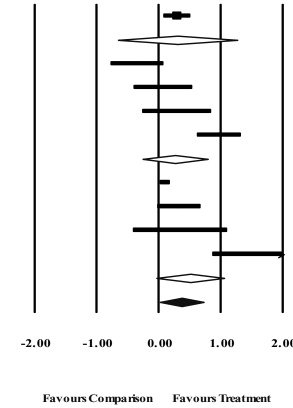
under randomeffects model

Figure 6. Post-secondary Achievement Grouped by Comparison Groups

Post-secondary Achievement Grouped by Comparison Groups

Group by Comparison	Study name	Outcome	Comparison	Statistics for each study: Hedges's g	Standard error	Lower limit	Upper limit	p-Value
mixed-age peers	Barnett & Durden, 1993	Post-secondary Achievement	mixed-age peers	0.294	0.114	0.071	0.517	0.010
mixed-age peers				0.294	0.490	-0.667	1.255	0.549
older age peers	Janos, Robinson, & Lunneborg, 1989	Post-secondary Achievement	older age peers	-0.350	0.221	-0.784	0.084	0.114
older age peers	Noble & Robinson, 1993	Post-secondary Achievement	older age peers	0.067	0.245	-0.413	0.547	0.784
older age peers	Janos & Robinson, 1985	Post-secondary Achievement	older age peers	0.287	0.286	-0.274	0.848	0.316
older age peers	Brody, Assouline, & Stanley, 1990	Post-secondary Achievement	older age peers	0.971	0.184	0.610	1.332	0.000
older age peers				0.255	0.266	-0.266	0.776	0.337
same age peers	Richards, 2006	Post-secondary Achievement	same age peers	0.099	0.045	0.011	0.187	0.027
same age peers	Weiner, 1985	Post-secondary Achievement	same age peers	0.324	0.182	-0.032	0.680	0.074
same age peers	Swiatek & Benbow, 1991	Post-secondary Achievement	same age peers	0.343	0.390	-0.421	1.107	0.379
same age peers	Janos, 1987	Post-secondary Achievement	same age peers	1.809	0.487	0.855	2.763	0.000
same age peers				0.498	0.277	-0.045	1.042	0.072
Overall				0.361	0.179	0.011	0.711	0.043

Hedges's g and 95% CI



under randomeffects model

Table 11. Summary of Combined Effect Sizes for Academic Achievement

Outcome group		No. of studies	Combined ES	Standard error(SE)	95% Confidence Intervals		Test of null (2-tail) p-value
					Lower limit	Upper limit	
Achievement overall		28	.180	.128	-.072	.431	.162
Achievement	P-12 achievement	21	.147	.164	-.174	.467	.370
	Post-secondary achievement	7	.313	.294	-.262	.889	.286
Achievement	with same age peers	13	.396	.187	.029	.762	.034*
	with older age peers	9	.224	.222	-.212	.660	.315
	with mixed-age peers	6	-.323	.265	-.842	.197	.223
P-12 achievement	with same age peers	9	.347	.261	-.165	.858	.184
	with older age peers	6	.334	.319	-.290	.959	.294
	with mixed-age peers	6	-.320	.314	-.936	.295	.308

Outcome group		No. of studies	Combined ES	Standard error(SE)	95% Confidence Intervals		Test of null (2-tail) p-value
					Lower limit	Upper limit	
Post-secondary achievement	with same age peers	4	.498	.277	-.045	1.042	.072
	with older age peers	4	.255	.266	-.266	.776	.337
	with mixed-age peers	1	.294	.490	-.667	1.255	.549

*$p < 0.05$ **$p < 0.01$ ***$p < 0.001$

5.1. 2 Heterogeneity Analysis

Table 12 presents a summary of the heterogeneity analysis results for academic achievement. As Table 12 shows, among the total twelve groups, there were nine groups of effect sizes that exhibited high degrees of heterogeneity. The heterogeneity tests of all of these nine groups each produced a result that was statistically significant at the significance level of $p < .01$. These nine groups were Achievement Overall, P-12 Achievement, Post-secondary Achievement, Achievement with Same Age Peers, Achievement with Mixed-Age Peers, P-12 Achievement with Same Age Peers, P-12 Achievement with Mixed-Age Peers, Post-secondary Achievement with Same Age Peers, and Post-secondary Achievement with Older Age Peers. Furthermore, the I-squared values for these nine groups were 95.786, 96.775, 67.934, 86.699, 98.786, 88.555, 98.786, 78.279, and 86.782, respectively, all indicating high degrees of heterogeneity.

Based on the heterogeneity analysis results above, further testing for moderators was performed on two groups: P-12 Achievement and Achievement with Same Age Peers. No further analyses were conducted on the rest of seven groups although they all showed high degrees of heterogeneities. There were two reasons for this decision: First, when using a method that is analogous to analysis of variance (ANOVA) to test for moderators, it requires at least 10 studies (effect sizes) for each group that is being analyzed (Higgins & Green, 2006). In this case, only three groups among the nine had sufficient studies for reliable testing of moderators. These three groups were Achievement Overall, P-12 Achievement, and Achievement with Same Age Peers, which consisted of 28, 21, and 13 studies (effect sizes), respectively.

Second, the P-12 Achievement group can be considered as an adequate representative for the group of Achievement Overall and it appears to be a better candidate group for testing for moderators than the latter. Specifically, P-12 Achievement accounted for the majority of the total variance when the achievement results were grouped into P-12 Achievement and Post-secondary Achievement. As Table 12 shows, with a Q-statistic value of 620.092, the P-12

Achievement group actually accounted for 96.8% of the total variances (Q = 640.7), whereas the Post-secondary Achievement group (Q = 18.711) only accounted for 3.2% of the total variance. Furthermore, compared to the group of Achievement Overall, P-12 Achievement can be considered to be a better candidate group for testing for moderators because it only contained the results for one outcome level (i.e., P-12 level).

In summary, for the reasons discussed above, further testing for moderators was performed on the P-12 Achievement and Achievement with Same Age Peers groups of studies.

Table 12. Summary of Heterogeneity Analysis for Academic Achievement

Outcome group/sub-group		No. of studies	Combined ES	Q-value	df (Q)	Heterogeneity p-value	I^2
Achievement overall		28	.180	640.700	27	.000***	95.786
Achievement	P-12 achievement	21	.147	620.092	20	.000***	96.775
	Post-secondary achievement	7	.313	18.711	6	.005 **	67.934
Achievement	with same age peers	13	.396	90.218	12	.000***	86.699
	with older age peers	9	.224	15.310	8	.053	47.746
	with mixed-age peers	6	-.323	411.967	5	.000***	98.786
P-12 achievement	with same age peers	9	.347	69.897	8	.000***	88.555
	with older age peers	6	.334	8.822	5	.116	43.322
	with mixed-age peers	6	-.320	411.967	5	.000***	98.786
Post-secondary achievement	with same age peers	4	.498	13.812	3	.003**	78.279
	with older age peers	4	.255	22.697	3	.000***	86.782
	with mixed-age peers	1	.294	.000	0	1.000	.000

$^{*}p < 0.05$ $^{**}p < 0.01$ $^{***}p < 0.001$

5.1.3 Testing for Moderators

P-12 Achievement. A total of 25 variables were tested for the group P-12 Achievement (see Table 13). The 25 variables were designated to describe either the general or methodological features of the primary studies. Coding categories for each variable were listed in numerical order, using the numerical codes as displayed in *the Study Substantive Features Coding Descriptors* (see Appendix B). For example, for the variable of study focus, a number of “1” was used to denote the coding category showing that a study’s focus was academic achievement, a number of “2” was given to represent the coding category showing that a study’s focus was social-emotional development, and a number of “3” was associated with the coding category showing a study’s focus includes both academic achievement and social-emotional development.

It must be emphasized that not all of the variables that appeared in *the Study Feature Coding Descriptors (see* Appendix B) were tested for moderators. Some variables were not tested because they were found to have too much missing information. These variables include mean age at acceleration, and mean age of treatment/comparison group at data collection. Some variables were not tested because all studies were coded as the same category, so there was only one sub-group for a variable. These variables include treatment/comparison group categories, comparison perspective, and statistical power. The variables of test type and test subject were not tested because some studies qualified for more than one coding category for one or both of these two variables. The variables of page number and coding time were not tested because they were designated to provide additional statistical information for coding.

As Table 13 shows, effects were heterogeneous within most of the variables, as indicated by high Q_w and I^2 values. Four variables were significantly associated with the total variance. They were acceleration duration (*g* .072, 95% CI -.574 to .718, 21 studies, $p < .01$), comparison group construction (*g* -.335, 95% CI -1.100 to .429, 21 studies, $p < .001$), statistical analysis (*g* -.509, 95% CI -1.380 to .362, 21 studies, $p < .001$), and study validity status (*g* -.595, 95% CI -

1.589 to .400, 21 studies, p < .001). Thus, these four variables were identified as moderators for the group P-12 Achievement.

Although it is tempting to overinterpret the results of testing for moderators, the four moderators identified above were significantly associated with the total variances of the effect sizes. In other words, it can be said that there will be a difference on the effects of acceleration on P-12 level high ability learners when studies or acceleration interventions have different characteristics that are denoted by these four moderators. Take acceleration duration, for example. There will be a difference in the effects of an acceleration intervention if the lengths of the acceleration intervention in the primary studies fall into different coding categories (i.e., either under 14 weeks, between 14 to 28 weeks, over 28 weeks, or non-specific time period) of the variable of acceleration duration. As Table 13 shows, for the variable of acceleration duration, the effect sizes (Hedge's *g*) corresponding with the four coding categories were as follows: .800 (95% CI -.078 to 1.679, 2 studies), -.933 (95% CI -1.566 to -.299, 3 studies), .294 (95% CI -.026 to .615, 12 studies), and .202 (95% CI -.350 to .755, 4 studies). When the acceleration intervention lasted 'under 14 weeks', the effects appeared the highest (*g* .800), which was followed by 'over 28 weeks' (*g* = .294) and 'non-specific time period' (*g* =.202); whereas the duration length of '14 to 28 weeks' appeared to be negatively associated with the effects (*g* -.933). However, it should be noted that there were only 3 studies [including the Hsu (2003) study] involved in the analysis of the category of '14 to 28 weeks'. Thus, it is possible that the analysis results were under strong influence of the Hsu (2003) study, which had a very low effect size (*g* = -2.493).

Likewise, as depicted in Table 13, in terms of the variable of comparison group construction, studies with the approach of matching IQ, achievement scores, age or grade-level appeared to yield the highest effects (*g* .341, 95% CI .167 to .514, 12 studies), followed by those using self-reported information to construct comparison group (*g* .250, 95% CI -.423 to .923, 1 study). However, the studies with ex post facto design were found to have the lowest effects (*g* -

2.493, 95% CI -3.040 to -1.946, 1 study). One may wonder that the value of -2.493 seems unreasonably low. Two factors may help to explain this. First, it is noteworthy, as shown in Table 13, that there was only one study [i.e., Hsu (2003)] that fell into the coding category of ex post facto design, and, as noted above, this study had an extremely negative result. Second, it must be emphasized again that this meta-analysis, along with the testing for moderators, was performed under random effects model. Under random effects models, effects were combined with weighted procedures. As a result, extreme values may be more likely to appear when there are a limited number of studies (effect sizes) involved in the analysis.

For the variable of statistical analysis, studies conducted using student-level ANCOVA, MANCOVA, or logistic regression analysis were found to have the highest effects (*g* .363, 95% CI .147 to .580, 6 studies), closely followed by those using student-level t-test, ANOVA, or MANOVA (*g* .299, 95% CI .144 to .455, 12 studies). The lowest effects were found in studies using descriptive statistics (*g* -2.493, 95% CI -2.959 to -2.027, 1 study).

In terms of the variable of study validity status, studies with moderate validity were found to have the highest effects (*g* .308, 95% CI .084 to .533, 9 studies), followed by studies with high validity (*g* .210, 95% CI .014 to .407, 11 studies); whereas studies with low validity appeared to have the lowest effect (*g* -2.493, 95% CI -3.096 to -1.890, 1 study), as might be expected. Once again, it is noteworthy that under the case of this variable, only one study [i.e., Hsu (2003)] fell into the coding category that resulted in a low effect (*g* = -2.493). The low validity of this study suggests that its unusually negative effects on academic achievement should be interpreted with caution.

To address Research Question 4 "what differences exist between subject-based acceleration and grade-based acceleration in terms of their effects on high-ability learners", a significance test was conducted. The purpose of this significance test was to examine the difference between the following two sub-groups: one sub-group consisted of studies that investigated the effects of subject-based acceleration and another included studies that assessed the effects of grade-based acceleration. The test was performed after all the irrelevant studies were deleted. As shown in Table 13, there was one study examining both subject-based and grade-based acceleration, and two studies investigating some acceleration interventions that could not be categorized into neither subject-based nor grade-based acceleration. After these three studies were deleted, 18 studies remained in the analysis. No statistically significant difference between the effects of subject-based acceleration and grade-based acceleration was found (Q_b = .292, $df = 1$, $p > .05$).

Table 13. Results of Testing for Moderators on P-12 Achievement

Variable	No. of studies	*g*	Lower limit	Upper limit	Q_w(*df*)	I^2	Qb(*df*)	*p*
Study focus	**21**	**.150**	**-.202**	**.502**			**. 227(1)**	**.634**
1.Academic achievement	13	.083	-.365	.531	592.253(12)	97.974		
2.Social-emotional development								
3.Both	8	.262	-.307	.831	12.166(7)	42.461		
Identification	**21**	**.262**	**-.280**	**.804**			**3.762(7)**	**.807**
1.Achievement scores	9	-.185	-.798	.429	537.480(8)	98.512		
2.IQ								
3.Age	1	.652	-1.232	2.536	.000(0)	.000		
4.Teacher/Parent identification	1	.353	-1.485	2.191	.000(0)	.000		
5.Self-reported	2	.257	-1.035	1.549	4.847(1)	79.369		
6.Multiple ways	5	.287	-.544	1.117	7.362(4)	45.668		
7.Enrolled in gifted/enrichment program/class	1	1.671	-.400	3.742	.000(0)	.000		
8.Classified as G/T in regional, such as states, district, schools, etc	1	.442	-1.413	2.297	.000(0)	.000		
9.Qualified for certain class, school, college, or grade level	1	.206	-1.616	2.028	.000(0)	.000		
Acceleration category	**21**	**.152**	**-.221**	**.526**			**.345(3)**	**.951**
1. Grade-based acceleration	7	.283	-.364	.931	10.531(6)	43.024		
2. Subject-based acceleration	11	.050	-.467	.567	582.826(10)	98.287		
3. Both 1 and 2	1	.315	-1.377	2.007	.000(0)	.000		

Variable	No. of studies	*g*	Lower limit	Upper limit	$Q_w(df)$	I^2	Qb(*df*)	*p*
4. Non-specific	2	.174	-1.031	1.380	1.088(1)	8.093		
Acceleration forms	**21**	**.154**	**-.235**	**.543**			**.403(4)**	**.982**
1. Early kindergarten, school or college entrance	5	.254	-.544	1.053	4.418(4)	9.454		
2. Grade skipping								
3. Advanced placement/Dual Credits/ International Baccalaureate	1	.206	-1.544	1.956	0.000(0)	.000		
4. Subject-matter acceleration (e.g., math and reading)	8	-.005	-.639	.629	554.996(7)	98.739		
5. Curriculum compacting	1	.353	-1.413	2.119	0.000(0)	.000		
6. Multiple forms	6	.236	-.490	.962	7.776(5)	36.700		
7. Early graduation								
8. Mentoring								
Acceleration duration**	**21**	**.072**	**-.574**	**.718**			**14.138(3)**	**.003****
1.Under 14 weeks	2	.800	-.078	1.679	6.540(1)	84.709		
2.14 to 28 weeks	3	-.933	-1.566	-.299	207.964(2)	99.038		
3.Over 28 weeks	12	.294	-.026	.615	36.477(11)	68.844		
4.Non-specific	4	.202	-.350	.755	1.348(3)	.000		

Variable	No. of studies	*g*	Lower limit	Upper limit	$Q_w(df)$	I^2	Qb(*df*)	*p*
Grade level at acceleration	**21**	**.198**	**-.302**	**.698**			**3.603(6)**	**.730**
1. Elementary	7	.142	-.532	.816	59.665(6)	89.944		
2. Middle school	1	1.671	-.344	3.686	0.000(0)	.000		
3. High school	5	-.273	-1.065	.518	464.984(4)	99.140		
4. Post secondary	2	.324	-.946	1.595	.581(1)	.000		
5. Both 1 and 2	3	.342	-.687	1.371	5.735(2)	65.125		
6. Both 2 and 3	1	-.020	-1.785	1.745	0.000(0)	.000		
7. Both 3 and 4								
8. Kindergarten	2	.325	-.945	1.595	.027(1)	.000		
9. varying (include multiple levels)								
Study time	**21**	**.150**	**-.201**	**.501**			**.388(2)**	**.824**
1. Study conducted within 1 year after acceleration	8	.008	-.564	.580	563.951(7)	98.759		
2. Longitudinal studies	11	.229	-.254	.712	15.159(10)	34.031		
3. Both	2	.276	-.862	1.414	.033(1)	0.000		
School type	**21**	**.153**	**-.223**	**.528**			**.545(3)**	**.909**
1. Rural								
2. Suburban	4	.322	-.557	1.202	46.859(3)	93.598		
3. Urban	4	.223	-.636	1.083	1.188(3)	.000		

Variable	No. of studies	*g*	Lower limit	Upper limit	$Q_w(df)$	I^2	Qb(*df*)	*p*
4. Not applicable	10	.010	-.530	.551	552.149(9)	98.370		
5. National	3	.321	-.670	1.312	7.234(2)	72.353		
SES	**21**	**.154**	**-.226**	**.534**			**1.041(4)**	**.904**
1. Low	2	.758	-.540	2.056	8.513(1)	88.254		
2. Medium	4	.023	-.829	.876	30.378(3)	90.124		
3. High	1	.039	-1.685	1.763	0.000(0)	0.000		
4. Non-specific	12	.095	-.399	.589	565.336(11)	98.054		
5. Low to medium	2	.287	-.920	1.494	7.160(1)	86.033		
6. Medium to high								
Ethnicity	**21**	**.153**	**-.222**	**.527**			**.681(3)**	**.878**
1. Non-minority students dominant (over 60%)	4	.307	-.547	1.161	7.191(3)	58.284		
2. Minority students dominant (over 60%)	2	-.304	-1.515	.907	6.142(1)	83.719		
3. Approximately equivalent	1	.206	-1.475	1.887	0.000	0.000		
4. Non-specific	14	.169	-.291	.630	571.906	97.727		
Gender	**21**	**.155**	**-.290**	**.600**			**3.865(3)**	**.276**
1. Male predominant (over 60%)	5	.370	-.272	1.012	5.100(4)	21.575		
2. Female predominant (over 60%)	4	.207	-.514	.928	2.402(3)	0.000		
3. Approximately equivalent	6	.416	-.189	1.021	19.974(5)	74.967		

Variable	No. of studies	*g*	Lower limit	Upper limit	Q_w(*df*)	I^2	Qb(*df*)	*p*
4. Non-specific	6	-.339	-.928	.250	370.405(5)	98.650		
Study retrieval source	**21**	**.151**	**-.207**	**.509**			**.141(2)**	**.932**
1.Journal article	13	.136	-.319	.591	571.069(12)	97.899		
2.Doctoral dissertation	6	.236	-.432	.904	9.039(5)	44.687		
3.Unpublished study obtained from author	2	-.011	-1.174	1.152	18.862(1)	94.698		
Sample assignment	**21**	**.151**	**-.204**	**.505**			**.333(1)**	**.564**
1. Randomized	6	.320	-.356	.996	46.167(5)	89.170		
2. Non-randomized	15	.086	-.331	.503	572.983(14)	97.557		
3. Mixed								
Comparison group construction***	**21**	**-.335**	**-1.100**	**.429**			**93.835(4)**	**.000*****
1. Matching on IQ, achievement scores, age, or grade-level.	12	.341	.167	.514	25.475(11)	56.821		
2. Ex post facto design	1	-2.493	-3.040	-1.946	.000(0)	.000		
3. Randomization from initial pool of eligible subjects	5	.104	-.163	.370	41.350(4)	90.327		
4. Randomization from clustered sample groups								
5. Self-reported information	1	.250	-.423	.923	.000(0)	.000		

Variable	No. of studies	*g*	Lower limit	Upper limit	Q_w(*df*)	I^2	Qb(*df*)	*p*
6. The rest of eligible subjects after treated groups have been assigned.	2	.038	-.427	.502	.009(1)	.000		
Subject group numbers	**21**	**.150**	**-.202**	**.502**			**.218(1)**	**.641**
1. Two groups	15	.097	-.321	.515	590.299(14)	97.628		
2. More than two groups	6	.281	-.371	.933	11.366(5)	56.009		
Test design	**21**	**.132**	**-.208**	**.472**			**1.231(2)**	**.540**
1.Posttest only	13	.278	-.115	.671	21.698(12)	44.695		
2. Pretest-posttest	5	-.125	-.778	.528	354.778(4)	98.873		
3. Multiple times tests	3	-.005	-.819	.809	55.768(2)	96.414		
Research design	**21**	**.158**	**-.239**	**.556**			**1.247(2)**	**.536**
1. Experimental	3	-.083	-1.008	.842	23.729(2)	91.571		
2. Quasi-experimental	3	.633	-.340	1.605	9.200(2)	78.262		
3. Causal-comparative research	15	.110	-.303	.523	577.978(14)	97.578		
Sample size	**21**	**.142**	**-.371**	**.655**			**5.131(3)**	**.162**
1. ≤ 100	3	.727	-.107	1.560	5.692(2)	64.865		
2. 101-500	12	.158	-.228	.545	69.772(11)	84.234		
3. 501-1000	3	-.558	-1.334	.218	288.340(2)	99.306		
4. >1000	3	.271	-.492	1.033	5.470(2)	63.436		
5. Not applicable								

Variable	No. of studies	*g*	Lower limit	Upper limit	$Q_w(df)$	I^2	Qb(*df*)	*p*
Statistical analysis***	**21**	**-.509**	**-1.380**	**.362**			**136.507(3)**	**.000*****
1. Descriptive statistics (including dichotomous outcomes)	1	-2.493	-2.959	-2.027	.000(0)	.000		
2. Student-level T-test, ANOVA, or MANOVA	12	.299	.144	.455	31.292(11)	63.847		
3. Student-level ANCOVA, MANCOVA, Logistic Regression Analysis	6	.363	.147	.580	8.809(5)	43.238		
4. Cluster-level analysis								
5. Hierarchical Linear Model (HLM) analysis with cluster-level assignment	2	-.314	-.655	.026	10.908(1)	90.833		
Data source	**21**	**.159**	**-.206**	**.523**			**2.061(5)**	**.841**
1. Standardized tests	5	.324	-.335	.983	10.517(4)	61.967		
2. Non-standardized tests	6	-.228	-.843	.388	350.105(5)	98.572		
3. Survey/questionnaires	2	.263	-.794	1.320	.129(1)	.000		
4. Combing 1, or 2, or 3, or 5.	4	.218	-.520	.956	2.018(3)	.000		
5. Accumulative records or archival data	1	.436	-1.026	1.898	.000(0)	.000		
6. National databases	3	.302	-.550	1.155	7.032(2)	71.560		

Variable	No. of studies	*g*	Lower limit	Upper limit	$Q_w(df)$	I^2	Qb(*df*)	*p*
Reliability information	**21**	**.151**	**-.204**	**.505**			**.133(1)**	**.716**
1. Yes	17	.118	-.278	.513	599.209(16)	97.330		
2. No	4	.283	-.515	1.082	6.502(3)	53.863		
Measurement reliability status	**21**	**.147**	**-.179**	**.473**			**.663(3)**	**.882**
1. High reliability	11	.195	-.255	.644	72.198(10)	86.149		
2. Moderate reliability	5	-.096	-0.783	.591	336.354(4)	98.811		
3. Low reliability	1	.206	-1.249	1.661	.000(0)	.000		
4. Not-available	4	.283	-.449	1.016	6.502(3)	53.863		
Study validity status***	**21**	**-.595**	**-1.589**	**.400**			**75.454(2)**	**.000*****
1. Low validity	1	-2.493	-3.096	-1.890	.000(0)	.000		
2. Moderate validity	9	.308	.084	.533	12.543(8)	36.218		
3. High validity	11	.210	.014	.407	74.658(10)	86.606		
ES reporting	**21**	**.150**	**-.195**	**.494**			**.021(1)**	**.885**
1. Yes	9	.178	-.343	.700	66.254(8)	87.925		
2. No	12	.127	-.333	.587	522.307(11)	97.894		
Effect size extraction	**21**	**.279**	**-.307**	**.865**			**2.440(2)**	**.295**
1. Calculated from descriptive statistics	16	.001	-.388	.389	531.737(15)	97.179		

Variable	No. of studies	*g*	Lower limit	Upper limit	Q_w(*df*)	I^2	Qb(*df*)	*p*
2. Calculated from inferential statistics	4	.645	-.149	1.440	14.820(3)	79.758		
3. Estimated from probabilities								
4. Computed effect sizes in original study	1	.652	-.947	2.251	.000(0)	.000		

**p* < 0.05 ***p* < 0.01 ****p* < 0.001

Achievement with Same Age Peers. Likewise, a total of 25 variables were tested for moderators for all studies that investigated achievement of accelerants in comparison to their same age peers (see Table 14). As Table 14 indicates, effects were heterogeneous within most of the tested variables. Nine variables were significantly associated with the total variance. They were identification (*g* .544, 95% CI .096 to .992, 13 studies, p < .05), acceleration category (*g* .630, 95% CI -.111 to 1.370, 13 studies, p < .05), acceleration duration (*g* .353, 95% CI -.098 to .804, 13 studies, p< .01), grade level at acceleration (*g* .565, 95% CI .087 to 1.043, 13 studies, p < .05), SES (*g* .302, 95% CI -.237 to .842, 13 studies, p < .001), ethnicity (*g* .142, 95% CI -.307 to .590, 13 studies, p < .05), study retrieval source (*g* .103, 95% CI -.423 to .629, 13 studies, p < .001), statistical analysis (*g* .114, 95% CI -.404 to .633, 13 studies, p < .001), and effect size reporting (*g* .345, 95% CI -.100 to .791, 13 studies, p < .05). Thus, these nine variables were identified as moderators of the combined effects of investigations of the academic effects of accelerating high ability students when the accelerants are compared with similarly high ability students who are not accelerated.

Table 14 also shows the magnitude and the dispersion of the effect of each coding category within each tested variable. However, because there were only a total of 13 studies involved in the analysis and too often for some variables only one study that fell into a certain category, the practical implication of these results is quite limited. Results related to the following variables may be noteworthy because the study distributions within each variable were comparatively reasonable. For example, for the variable of acceleration duration, it can be seen that when the acceleration intervention lasted 'under 14 weeks' the effects appeared the highest (*g* 1.671, 95% CI .585 to 2.757, 1 study), which were followed by 'over 28 weeks' (*g* .531, 95% CI .294 to .767, 6 studies) and 'non-specific time period' (*g* .145, 95% CI -.106 to .397, 4 studies); whereas the duration length of '14 to 28 weeks' appeared to be negatively associated with the effects (*g* -.168, 95% CI -.545 to .210, 2 studies). Comparing the results for acceleration duration in this group with the results in the group of P-12 Achievement, it can be seen that for both groups when the acceleration intervention lasts ‘over 28 weeks’, comparatively larger effects appear, whereas the duration length of ’14 to 28 weeks’ is negatively associated with the effects.

For the variable of study retrieval source, it is interesting to note that ‘journal article’ exhibited higher effects (*g* .587, 95% CI .356 to .818, 6 studies) than did ‘doctoral dissertation’ (*g* .241, 95% CI .059 to .422, 6 studies). For another methodological feature variable---statistical analysis, the results showed that the studies with ‘student-level t-test, ANOVA, or MANOVA’ (*g* .569, 95% CI .315 to .823, 6 studies) yielded higher effect than did those with ‘student-level ANCOVA, MANCOVA, or logistic regression analysis’ (*g* .273, 95% CI .082 to .465, 6 studies).

For the variable of effect size reporting, results showed that the combined effect size of the studies in which researchers provided no direct effect size information were higher (*g* .124, 95% CI -.152 to .400, 7 studies) than those in which researchers reported direct effect size information (*g* .579, 95% CI .268 to .889, 6 studies).

Again, to determine if any difference existed between subject-based acceleration and grade-based acceleration in terms of their effects on academic achievement when accelerants are compared with their same age peers, a significant test was conducted. As shown in Table 14, there was no study examining both subject-based and grade-based acceleration, and there was one study investigating an acceleration intervention that could not be categorized into either subject-based or grade-based acceleration. Thus, 12 studies remained in the test. Statistically significant differences between subject-based acceleration and grade-based acceleration were seen for the combined effects ($Qb = 6.786$, $df = 1$, $p < .01$). Specifically, the results showed that subject-based acceleration appeared to have a higher impact (*g*. 1.809, 95% CI .671 to 2.947, 11 studies) than did grade-based acceleration (*g* .269, 95% CI -.562 to 2.429, 1 study). However, the analysis may not be valid, as there was a substantial imbalance in the number of studies in the two groups and only one study in the grade-based acceleration group.

Table 14. Results of Testing for Moderators on Achievement with Same Age Peers

Variable	No. of studies	*g*	Lower limit	Upper limit	Q_w (df)	I^2	Q_b(*df*)	*p*
Study focus	**13**	**.393**	**.025**	**.760**			**2.620(2)**	**.270**
1.Academic achievement	9	.236	-.004	.475	71.829(8)	88.862		
2.Social-emotional development	1	.343	-.650	1.335	.000(0)	.000		
3.Both	3	.674	-.200	.679	8.434(2)	76.285		
Identification*	**13**	**.544**	**.096**	**.992**			**12.760(6)**	**.047***
1.Achievement scores	5	.111	-.242	.465	56.194(4)	92.882		
2.IQ								
3.Age								
4.Teacher/Parent identification	1	.353	-.423	1.129	.000(0)	.000		
5.Self-reported	1	1.809	.616	2.971	.000(0)	.000		
6.Multiple ways	2	.505	-.118	1.588	.360(1)	.000		
7.Enrolled in gifted/enrichment program/class	1	1.671	.440	2.660	.000(0)	.000		
8.Classified as G/T in regional, such as states, district, schools, etc	1	.442	-.375	1.061	.000(0)	.000		
9.Qualified for certain class, school, college, or grade level	2	.151	-.365	.667	1.145(1)	12.656		
Acceleration category	**13**	**.630**	**-.111**	**1.370**			**6.786(2)**	**.034***
1. Grade-based acceleration	1	1.809	.671	2.947	.000(0)	.000		
2. Subject-based acceleration	11	.269	.050	.487	78.782(10)	87.307		
3. Both 1 and 2								

Variable	No. of studies	*g*	Lower limit	Upper limit	Q_w (df)	I^2	Q_b(*df*)	*p*
4. Non-specific	1	.324	-.391	1.039	.000(0)	.000		
Acceleration forms	**13**	**.387**	**-.021**	**.794**			**6.502(5)**	**.260**
1. Early kindergarten, school or college entrance	1	1.809	.589	3.029	.000(0)	.000		
2. Grade skipping								
3. Advanced placement/Dual Credits/ International Baccalaureate	2	.151	-.395	.698	1.145(1)	12.656		
4. Subject-matter acceleration (e.g., math and reading)	7	.356	.019	.692	66.778(6)	91.015		
5. Curriculum compacting	1	.353	-.464	1.170	.000(0)	.000		
6. Multiple forms	1	.039	-.821	.899	.000(0)	.000		
7. Early graduation								
8. Mentoring	1	.324	-.516	1.164	.000(0)	.000		
Acceleration duration**	**13**	**.353**	**-.098**	**.804**			**17.101(3)**	**.001****
1.Under 14 weeks	1	1.671	.585	2.757	.000(0)	.000		
2.14 to 28 weeks	2	-.168	-.545	.210	19.441(1)	94.856		
3.Over 28 weeks	6	.531	.294	.767	10.143(5)	50.707		
4.Non-specific	4	.145	-.106	.397	1.619(3)	.000		

Variable	No. of studies	*g*	Lower limit	Upper limit	Q_w (df)	I^2	Q_b(*df*)	*p*
Grade level at acceleration*	**13**	**.565**	**.087**	**1.043**			**11.733(5)**	**.039***
1. Elementary	5	.194	-.142	.530	56.160(4)	92.878		
2. Middle school	1	1.671	.461	2.881	.000(0)	.000		
3. High school	4	.234	-.130	.597	10.552(3)	71.570		
4. Post secondary	1	1.809	.637	2.981	.000(0)	.000		
5. Both 1 and 2								
6. Both 2 and 3	1	.343	-.680	1.366	.000(0)	.000		
7. Both 3 and 4								
8. Kindergarten								
9. varying (include multiple levels)	1	.324	-.444	1.092	.000(0)	.000		
Study time	**13**	**.314**	**.106**	**.575**			**.620(2)**	**.733**
1. Study conducted within 1 year after acceleration	8	.288	.000	.577	78.530(7)	91.086		
2. Longitudinal studies	4	.496	.042	.949	10.627(3)	71.770		
3. Both	1	.250	-.610	1.110	.000(0)	.000		
School type	**13**	**.352**	**.097**	**.606**			**.254(3)**	**.968**
1. Rural								
2. Suburban	3	.284	-.265	.833	46.854(2)	95.731		
3. Urban	1	.206	-.623	1.035	.000(0)	.000		
4. Not applicable	8	.398	.070	.726	37.629(7)	81.397		

Variable	No. of studies	*g*	Lower limit	Upper limit	Q_w (df)	I^2	$Q_b(df)$	*p*
5. National	1	.353	-.510	1.216	.000(0)	.000		
SES*	**13**	**.302**	**-.237**	**.842**			**20.553(4)**	**.000***
1. Low	1	1.671	.575	2.767	.000(0)	.000		
2. Medium	1	-.629	-1.150	-.108	.000(0)	.000		
3. High	1	.039	-.560	.638	.000(0)	.000		
4. Non-specific	9	.363	.173	.553	37.536(8)	78.687		
5. Low to medium	1	.594	.059	1.129	.000(0)	.000		
6. Medium to high								
Ethnicity*	**13**	**.142**	**-.307**	**.590**			**11.422(3)**	**.010***
1. Non-minority students dominant (over 60%)	2	.317	-.088	.721	9.801(1)	89.797		
2. Minority students dominant (over 60%)	1	-.629	-1.227	-.031	.000(0)	.000		
3. Approximately equivalent	1	.206	-.356	.768	.000(0)	.000		
4. Non-specific	9	.474	.236	.712	20.449(8)	62.144		
Gender	**13**	**.368**	**-.007**	**.742**			**6.499(3)**	**.090**
1. Male predominant (over 60%)	3	.750	.238	1.262	6.746(2)	70.355		
2. Female predominant (over 60%)	2	.260	-.251	.771	.340(1)	.000		
3. Approximately equivalent	4	.522	.109	.935	10.628(3)	71.772		
4. Non-specific	4	.007	-.358	.372	30.159(3)	90.053		

Variable	No. of studies	*g*	Lower limit	Upper limit	Q_w (df)	I^2	Q_b(*df*)	*p*
Study retrieval source*	**13**	**.103**	**-.423**	**.629**			**23.581(2)**	**.000***
1.Journal article	6	.587	.356	.818	13.803(5)	63.777		
2.Doctoral dissertation	6	.241	.059	.422	11.625(5)	56.989		
3.Unpublished study obtained	1	-.629	-1.066	-.192	.000(0)	.000		
Sample assignment	**13**	**.335**	**.109**	**.561**			**1.005(1)**	**.316**
1. Randomized	4	.164	-.239	.567	41.323(3)	92.740		
2. Non-randomized	9	.413	.142	.684	43.315(8)	81.531		
3. Mixed								
Comparison group construction	**13**	**.315**	**-.009**	**.639**			**3.482(4)**	**.481**
1. Matching on IQ, achievement scores, age, or grade-level.	5	.651	.246	1.056	7.246(4)	44.799		
2. Ex post facto design	1	.099	-.662	.860	.000(0)	.000		
3. Randomization from initial pool of eligible subjects	4	.179	-.252	.611	41.323(3)	92.740		
4. Randomization from clustered sample groups								
5. Self-reported information	1	.250	-.630	1.130	.000(0)	.000		
6. The rest of eligible subjects after treated groups have been assigned.	2	.185	-.413	.793	1.083(1)	7.664		

Variable	No. of studies	*g*	Lower limit	Upper limit	Q_w (df)	I^2	Q_b(*df*)	*p*
Subject group numbers	**13**	**.346**	**.102**	**.590**			**.002(1)**	**.968**
1. Two groups	10	.349	.063	.636	75.305(9)	88.049		
2. More than two groups	3	.338	-.128	.804	11.652(2)	82.835		
Test design	**13**	**.330**	**.026**	**.635**			**1.899(2)**	**.387**
1.Posttest only	7	.407	.072	.743	22.866(6)	73.760		
2. Pretest-posttest	4	.464	.006	.922	9.201(3)	67.394		
3. Multiple times tests	2	-.010	-.576	.557	55.016(1)	98.182		
Research design	**13**	**.273**	**-.110**	**.657**			**4.417(2)**	**.110**
1. Experimental	2	-.146	-.634	.342	22.960(1)	95.645		
2. Quasi-experimental	3	.478	.003	.959	9.200(2)	78.262		
3. Causal-comparative research	8	.415	.153	.677	41.474(7)	83.122		
Sample size	**13**	**.413**	**-.110**	**.935**			**5.261(2)**	**.072**
1. ≤ 100	4	.900	.364	1.436	11.300(3)	73.452		
2. 101-500	8	.216	-.078	.509	61.880(7)	88.688		
3. 501-1000								
4. >1000	1	.099	-.675	.873	.000(0)	.000		
5. Not applicable								
Statistical analysis* **	**13**	**.114**	**-.404**	**.633**			**19.462(2)**	**.000*****
1. Descriptive statistics (including dichotomous outcomes)								

Variable	No. of studies	*g*	Lower limit	Upper limit	Q_w (df)	I^2	Q_b(*df*)	*p*
2. Student-level T-test, ANOVA, or MANOVA	6	.569	.315	.823	15.056(5)	66.790		
3. Student-level ANCOVA, MANCOVA, Logistic Regression Analysis	6	.273	.082	.465	14.212(5)	64.818		
4. Cluster-level analysis								
5. Hierarchical Linear Model (HLM) analysis with cluster-level assignment	1	-.629	-1.097	-.161	.000(0)	.000		
Data source	**13**	**.342**	**.064**	**.621**			**2.334(3)**	**.506**
1. Standardized tests	4	.403	.020	.786	6.971(3)	56.963		
2. Non-standardized tests	4	.150	-.261	.560	39.336(3)	92.373		
3. Survey/questionnaires	3	.655	.099	1.212	8.313(2)	75.943		
4. Combing 1, or 2, or 3, or 5.	2	.225	-.324	.774	.032(1)	.000		
5. Accumulative records or archival data								
6. National databases								
Reliability information	**13**	**.349**	**.099**	**.599**			**.618(1)**	**.432**
1. Yes	11	.309	.040	.578	69.953(10)	85.705		
2. No	2	.598	-.071	1.268	12.235(1)	91.827		

Variable	No. of studies	*g*	Lower limit	Upper limit	Q_w (df)	I^2	Q_b(*df*)	*p*
Measurement reliability status	**13**	**.402**	**-.033**	**.838**			**2.665(2)**	**.264**
1. High reliability	8	.227	-.123	.578	59.451(7)	61.177		
2. Moderate reliability	4	.757	.208	1.306	16.372(3)	23.855		
3. Low reliability	1	.206	-.725	1.137	.000(0)	.000		
4. Not-available								
Study validity status	**13**	**.372**	**-.014**	**.759**			**2.263(2)**	**.323**
1. Low validity	1	.099	-.794	.992	.000(0)	.000		
2. Moderate validity	5	.671	.177	1.164	17.970(4)	77.740		
3. High validity	7	.242	-.114	.599	61.615(6)	90.262		
ES reporting*	**13**	**.345**	**-.100**	**.791**			**4.600(1)**	**.032***
1. Yes	6	.124	-.152	.400	60.872(5)	91.786		
2. No	7	.579	.268	.889	16.012(6)	62.528		
Effect size extraction	**13**	**.363**	**.007**	**.719**			**2.340(1)**	**.126**
1. Calculated from descriptive statistics	8	.201	-.084	.486	46.078(7)	84.808		
2. Calculated from inferential statistics	5	.566	.195	.937	43.266(4)	90.755		
3. Estimated from probabilities								
4. Computed effect sizes in original study								

*$p < 0.05$ **$p < 0.01$ ***$p < 0.001$

5.1.4 Cumulative Analysis

Figure 7 presents the results of cumulative analysis of the combined effect sizes for academic achievement sorted by year of study. These results were accumulated over successive studies. For example, the second row gave the analysis summary for the first two studies (as shown in Figure 7, “Brody, 1985” and “Janos & Robinson, 1985”), and the third row provided the analysis summary for the first three studies (i.e., “Brody, 1985”, “Janos & Robinson, 1985”, and “Weiner, 1985”). As the studies were sorted by year of study, these results illustrated the emerging change of the combined effects over time when each new study appeared.

As Figure 7 displays, an obvious pattern was the accumulated effect sizes for academic achievement were consistently higher through the periods of 1990s than those in the 2000s. Specifically, from the year of 1990 (i.e., Brody, Assouline, & Stanley, 1990) to 1999 (i.e., Washington, 1999), a total of 12 studies appeared successively. As each new study was involved, the accumulated effect was found to be positive and statistically significant at the significance levels of $p<.05$, $p<.01$, or $p<.001$. This was indicated by the fact that the confidence interval corresponding to each of the accumulated effect size did not cover the value of zero. Also, as can be seen in Figure 7, the confidence interval lines associated with these 12 accumulated effect sizes lay on the right side of the zero line, indicating that the effects of acceleration favored the treatment group during this period. As shown in the bottom part of Figure 7, from the year of 2001 (i.e., Chilton, 2001) to 2008 (i.e., Wells, Lohman, & Marron, 2008), 10 studies appeared successively during this period. With the exception of two cases [when the study of Chilton (2001), Moon & Callahan (2001) was involved, respectively], all the other accumulated effects were not statistically significant.

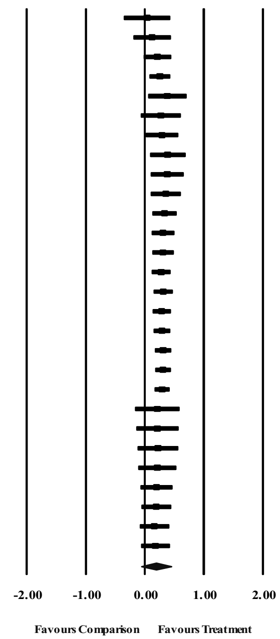
There are at least two possible ways to explain the observed change through the 1990s to the 2000s. First, it is possible that the educational environment or acceleration practices have changed since the year 2000 and the effects of acceleration have been negatively influenced by these changes. As a result, the effects of acceleration assessed through the primary studies have showed a decreasing pattern in the last decade. Second, in general, acceleration research has been conducted with more rigorous methods in recent year and this might lead to the comparatively smaller effect sizes reported in primary studies

Figure 7. Cumulative Analysis of Academic Achievement Sorted by Year of Study

Cumulative Analysis of Academic Achievement Sorted by Year of Study

Study name	Outcome	Comparison	Time point	Point	Standard error	Lower limit	Upper limit	p-Value
				Cumulative statistics				
Brody, 1985	P-12 Achievement	same age peers	1985	0.039	0.205	-0.363	0.441	0.849
Janos & Robinson, 1985	Post-secondary Achievement	older age peers	1985	0.123	0.167	-0.204	0.450	0.460
Weiner, 1985	Post-secondary Achievement	same age peers	1985	0.215	0.123	-0.026	0.456	0.080
Brody & Benbow, 1987	P-12 Achievement	mixed-age peers	1987	0.257	0.094	0.073	0.440	0.006
Janos, 1987	Post-secondary Achievement	same age peers	1987	0.383	0.169	0.051	0.715	0.024
Janos, Robinson, & Lunneborg, 1989	Post-secondary Achievement	older age peers	1989	0.273	0.178	-0.075	0.621	0.124
Brody, Assouline, & Stanley, 1990	P-12 Achievement	older age peers	1990	0.291	0.145	0.007	0.575	0.045
Callahan & Smith, 1990	P-12 Achievement	same age peers	1990	0.389	0.158	0.079	0.699	0.014
Swiatek & Benbow, 1991	Post-secondary Achievement	same age peers	1991	0.380	0.147	0.092	0.667	0.010
Swiatek, & Benbow, 1991 (2)	P-12 Achievement	older age peers	1991	0.355	0.133	0.094	0.616	0.008
Barnett & Durden, 1993	P-12 Achievement	mixed-age peers	1993	0.334	0.111	0.117	0.551	0.003
Noble & Robinson, 1993	Post-secondary Achievement	older age peers	1993	0.309	0.103	0.107	0.512	0.003
Reynolds, 1993	P-12 Achievement	older age peers	1993	0.309	0.096	0.121	0.497	0.001
Sayley & Brookshire, 1993	P-12 Achievement	older age peers	1993	0.277	0.085	0.109	0.444	0.001
Mills, Ablard, & Gustin, 1994	P-12 Achievement	same age peers	1994	0.312	0.086	0.144	0.480	0.000
Ambruster, 1995	P-12 Achievement	mixed-age peers	1995	0.288	0.082	0.126	0.449	0.000
Reis, Westberg, Kulikowich, & Purcell, 1998	P-12 Achievement	same age peers	1998	0.291	0.077	0.140	0.441	0.000
Washington, 1999	P-12 Achievement	same age peers	1999	0.312	0.074	0.166	0.457	0.000
Chilton, 2001	P-12 Achievement	same age peers	2001	0.308	0.071	0.169	0.447	0.000
Moon, & Callahan, 2001	P-12 Achievement	mixed-age peers	2001	0.296	0.069	0.160	0.431	0.000
Hsu, 2003	P-12 Achievement	mixed-age peers	2003	0.212	0.196	-0.172	0.596	0.279
Gagné & Gagnier, 2004	P-12 Achievement	older age peers	2004	0.215	0.187	-0.151	0.581	0.250
Ysseldyke, Tardrew, Betts, Thill, & Hannigan, 2004	P-12 Achievement	same age peers	2004	0.224	0.180	-0.129	0.578	0.213
Ma, 2005	P-12 Achievement	mixed-age peers	2005	0.212	0.170	-0.121	0.545	0.213
Richards, 2006	Post-secondary Achievement	same age peers	2006	0.198	0.143	-0.082	0.477	0.165
Fowler, 2007	P-12 Achievement	same age peers	2007	0.195	0.134	-0.066	0.457	0.143
Reis, Eckert, McCoach, Jacobs, & Coyne, 2008	P-12 Achievement	same age peers	2008	0.163	0.131	-0.094	0.420	0.213
Wells, Lohman, & Marron, 2008	P-12 Achievement	older age peers	2008	0.180	0.128	-0.072	0.431	0.162
				0.180	0.128	-0.072	0.431	0.162

Cumulative hedges's g (95% CI)



under random effects model

5.1.5 Assessment of Publication Bias

The assessment of publication bias for the meta-analysis of academic achievement was performed on the 28 studies included in this group. Among the 28 studies, 18 (64.3%) were published as journal articles, 8 (28.6%) were doctoral dissertations, and 2 (7.1%) were current studies obtained from researchers. As described in the methods section, two approaches were implemented to detect and adjust for the presence of publication bias. First, a funnel plot was produced to visually display the results of assessment of publication bias (see Figure 8).

In Figure 8, Hedges's *g* was plotted against its precision (i.e., 1/standard error), with the former along the x-axis and the latter along the y-axis. In general, larger studies with higher precision (smaller errors) tend to appear at the top of the funnel and cluster near the mean effect size; whereas smaller studies with lower precision (larger errors) tend to appear in the bottom with a disbursed distribution. If more studies (usually small studies) asymmetrically appear in the bottom of the funnel, there might be publication bias.

As Figure 8 shows, the majority of the studies asymmetrically clustered in the bottom part of the funnel. This suggests that there is presence of publication bias. It also reflects a situation that small studies were a major component of this meta-analysis of academic achievement. As a result, an existing perception was confirmed, which stated that acceleration research has often been conducted with small sample sizes in the field of gifted/talented education. However, visual inspection of funnel plots is generic and subjective. Further analysis was needed. Therefore, Duvall and Tweedie's trim and fill procedure was conducted, as previously described in the methods section.

Figure 8. Funnel Plot of Precision by Hedges's *g* (Academic Achievement)

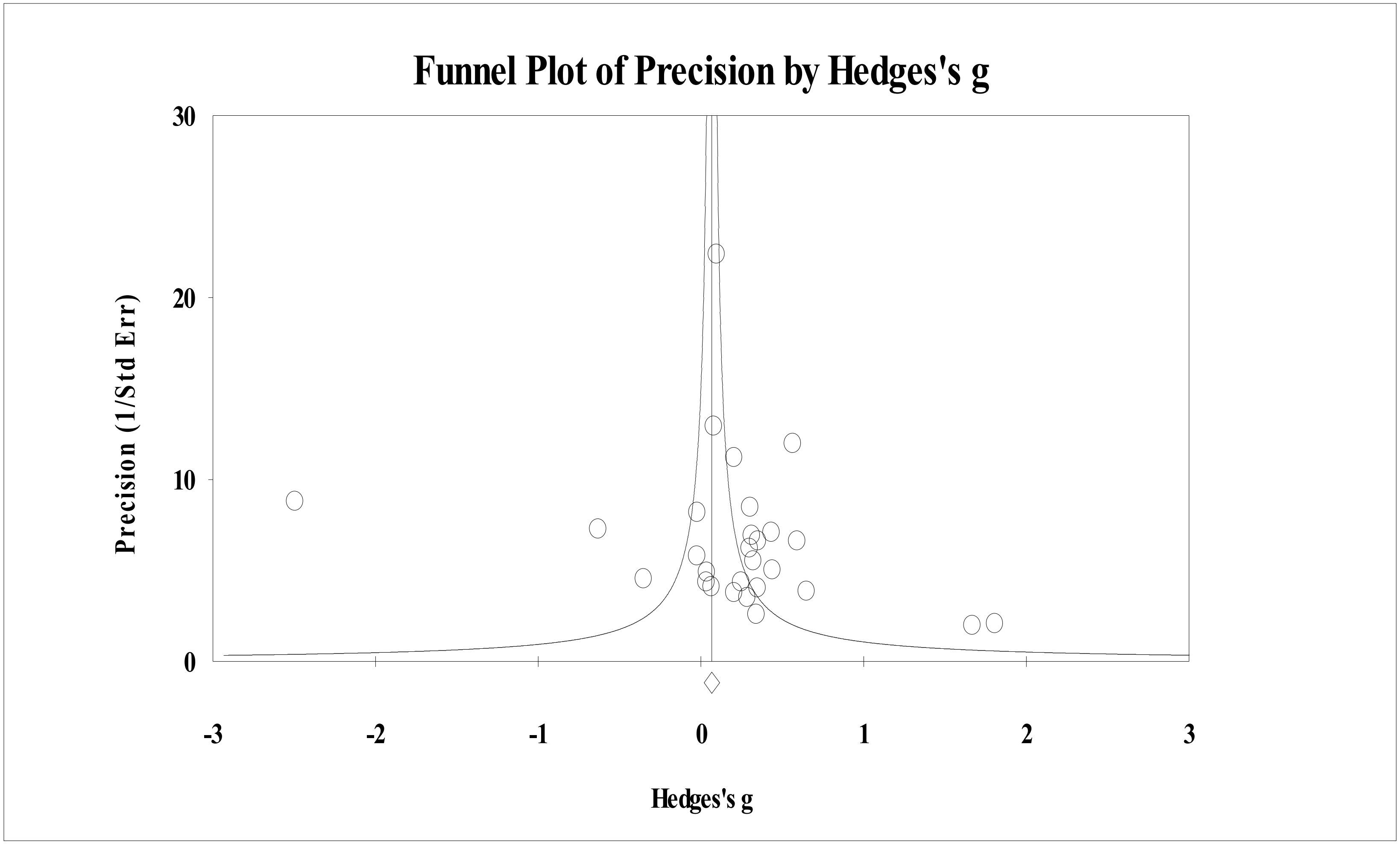
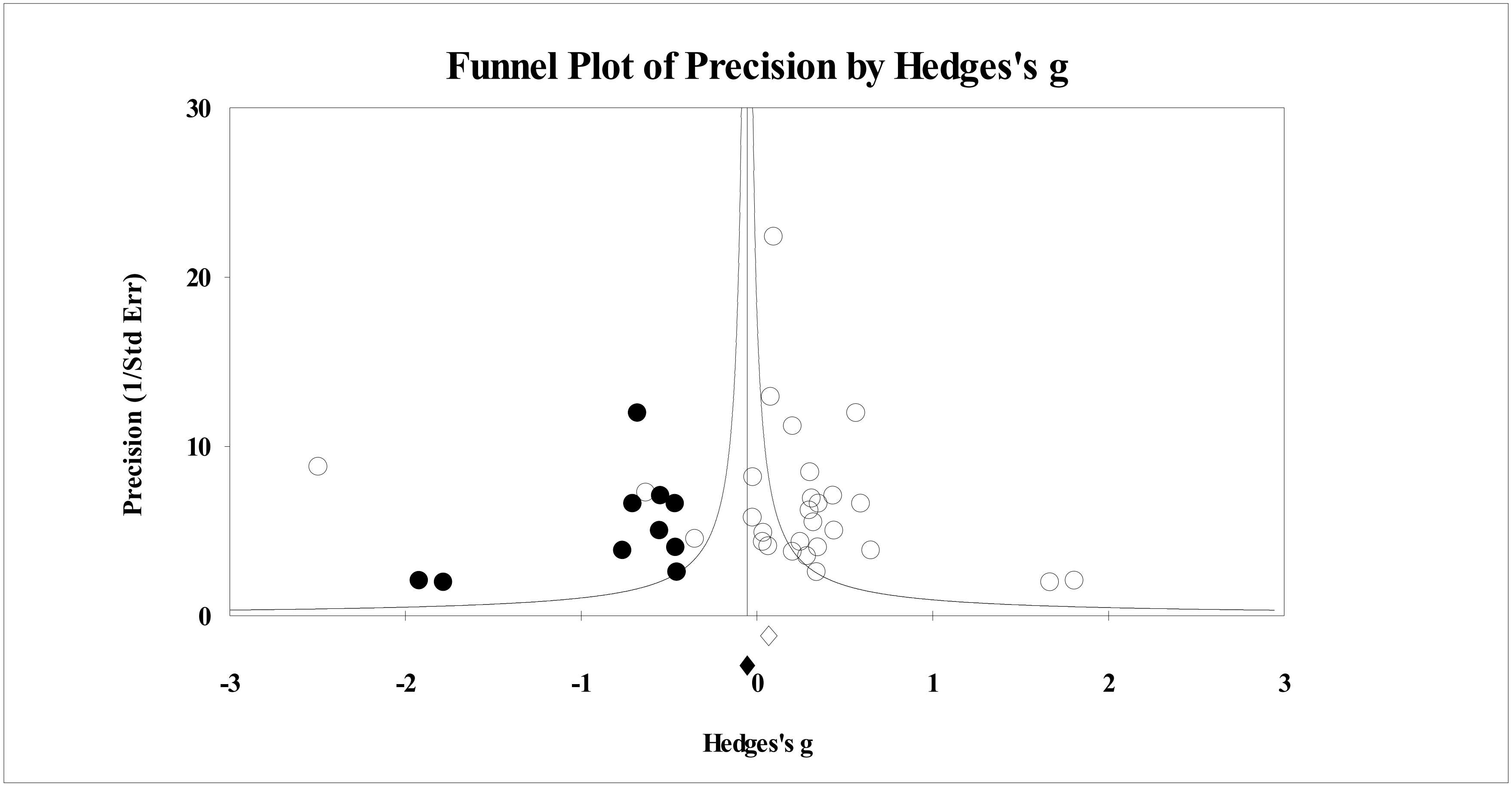


Figure 9 presents a funnel plot after Duvall and Tweedie's trim and fill procedure. In this funnel plot, black dots, an open diamond, and a closed diamond were used to denote the projected missing studies, the original (previously produced in the meta-analysis) and the recalculated (adjusted) combined effect size, respectively. As Figure 9 shows, 10 studies were projected to be missing from the left side of the mean effect line. This may still suggest the presence of publication bias, as did the visual inspection of the funnel plot in Figure 8. A comparison of the original combined effect size with the adjusted combined effect after the trim and fill suggested that the original combined effect size may be overestimated. In fact, statistics yielded with the trim and fill procedure showed that, under the random effects model, the point estimate and the 95% confidence interval for the combined effects was .180 (-.072, .431), while after the trim and fill procedure, the combined effects became -.059 (-.276, .159).

The results above appear informative. However, caution must be exercised in order to appropriately interpret them. For example, it should be noted that the presence of publication bias may be caused by lack of studies that reported negative effects of acceleration, rather than by lack of studies that were conducted with bigger sample sizes. As Figure 9 shows, all of the 10 projected missing studies appeared in the bottom left part of the funnel plot and no projected missing study appeared on the top of the funnel plot. It is reasonable to presume that if the publication bias is caused by the small sample size effects, then some projected missing studies might appear on the top after trim and fill adjustment. At the very least, it can be concluded that small sample size effects do not account for publication bias alone in this case.

Figure 9. Funnel Plot after Trim and Fill Procedure (Academic Achievement)



5.2 Social-Emotional Development Effects Results

5.2.1 Combined Effects

A separate meta-analysis was conducted on the effects of acceleration on social-emotional development. Figures 10 through 11 present the effect size information and forest plots for the social-emotional development effects. Table 15 summarizes the major statistical information on the combined effects that are presented in Figures 10 and 11. As Table 15 shows, the overall combined (average) effect size for social-emotional development was .076 (95% CI -.025 to .176, 22 studies). The significance test of the results showed that the effect was not statistically significant (df =21, p > .05).

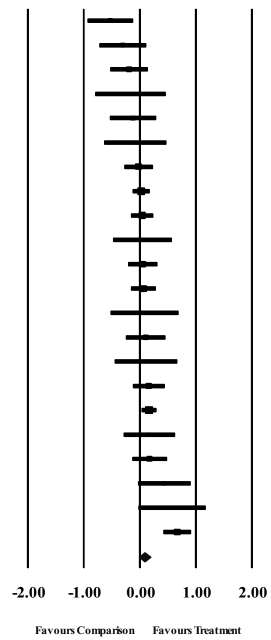
When the effect sizes for social-emotional development were grouped by comparison groups, three sub-groups were created: Social-Emotional Development with Older Age Peers, Social-Emotional Development with Same Age Peers, and Social-Emotional Development with Mixed-Age Peers. For the group of Social-Emotional Development with Older Age Peers, the combined (average) effect size was .052 (95% CI -.111 to .215, 10 studies). For the group of Social-Emotional Development with Same Age Peers, the combined (average) effect size was.141 (95% CI -.013 to .295, 9 studies). For the group of Social-Emotional Development with Mixed-Age Peers, the combined (average) effect size was -.036 (95% CI -.280 to .208, 3 studies). None of these results were statistically significant (df =9, df =8, df =2, respectively, p > .05).

Figure 10. Overall Combined Effect Size for Social-Emotional Development

Overall Combined Effect Size for Social-Emotional Development

Study name	Comparison	Hedges's g	Standard error	Lower limit	Upper limit	p-Value
		Statistics for each study				
Swiatek & Benbow, 1991	same age peers	-0.528	0.212	-0.944	-0.112	0.013
Sayler, 2008	older age peers	-0.306	0.219	-0.735	0.123	0.163
Ambruster, 1995	mixed-age peers	-0.193	0.176	-0.538	0.152	0.273
Ingersoll, 1992	older age peers	-0.171	0.326	-0.809	0.467	0.599
Janos, Robinson, & Lunneborg, 1989	older age peers	-0.124	0.214	-0.544	0.296	0.563
Robinson & Janos, 1986	older age peers	-0.084	0.288	-0.649	0.481	0.771
Brody & Benbow, 1987	mixed-age peers	-0.023	0.134	-0.286	0.240	0.864
Sayley & Brookshire, 1993	older age peers	0.024	0.084	-0.140	0.188	0.774
Plucker, & Taylor, 1998	same age peers	0.041	0.105	-0.165	0.247	0.696
Noble & Robinson, 1993	older age peers	0.046	0.272	-0.487	0.579	0.866
Swiatek, & Benbow, 1991 (2)	older age peers	0.053	0.138	-0.217	0.323	0.701
Ma, 2002	mixed-age peers	0.063	0.118	-0.169	0.295	0.594
Houston, 1999	same age peers	0.083	0.315	-0.534	0.699	0.793
Chilton, 2001	same age peers	0.105	0.184	-0.256	0.466	0.570
Cornell, Callahan, & Loyd, 1991	same age peers	0.110	0.290	-0.458	0.678	0.704
Washington, 1999	same age peers	0.160	0.148	-0.131	0.451	0.281
Lupkowski, Whitmore, & Ramsay, 1992	same age peers	0.163	0.071	0.024	0.302	0.021
Stamps, 2004	same age peers	0.168	0.239	-0.300	0.636	0.482
Gagné & Gagnier, 2004	older age peers	0.174	0.161	-0.142	0.490	0.281
Reynolds, 1993	older age peers	0.437	0.245	-0.043	0.917	0.074
Janos & Robinson, 1985	older age peers	0.574	0.311	-0.036	1.184	0.065
Jin & Moon, 2006	same age peers	0.664	0.130	0.408	0.920	0.000
		0.076	0.051	-0.025	0.176	0.139

Hedges's g and 95% CI



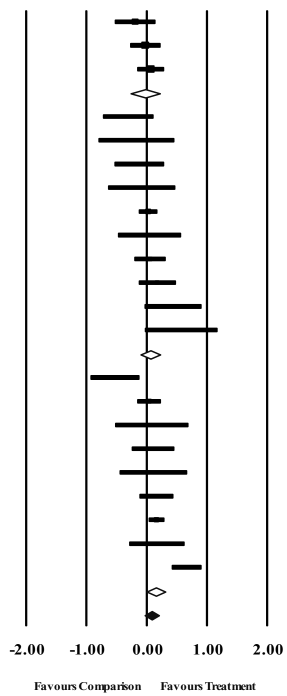
under random effects model

Figure 11. Social-Emotional Development Grouped by Comparison Groups

Social-Emotional Development Grouped by Comparison Groups

Group by Comparison	Study name	Comparison	Statistics for each study: Hedges's g	Standard error	Lower limit	Upper limit	p-Value
mixed-age peers	Ambruster, 1995	mixed-age peers	-0.193	0.176	-0.538	0.152	0.273
mixed-age peers	Brody & Benbow, 1987	mixed-age peers	-0.023	0.134	-0.286	0.240	0.864
mixed-age peers	Ma, 2002	mixed-age peers	0.063	0.118	-0.169	0.295	0.594
mixed-age peers			-0.036	0.124	-0.280	0.208	0.773
older age peers	Sayler, 2008	older age peers	-0.306	0.219	-0.735	0.123	0.163
older age peers	Ingersoll, 1992	older age peers	-0.171	0.326	-0.809	0.467	0.599
older age peers	Janos, Robinson, & Lunneborg, 1989	older age peers	-0.124	0.214	-0.544	0.296	0.563
older age peers	Robinson & Janos, 1986	older age peers	-0.084	0.288	-0.649	0.481	0.771
older age peers	Sayley & Brookshire, 1993	older age peers	0.024	0.084	-0.140	0.188	0.774
older age peers	Noble & Robinson, 1993	older age peers	0.046	0.272	-0.487	0.579	0.866
older age peers	Swiatek, & Benbow, 1991 (2)	older age peers	0.053	0.138	-0.217	0.323	0.701
older age peers	Gagné & Gagnier, 2004	older age peers	0.174	0.161	-0.142	0.490	0.281
older age peers	Reynolds, 1993	older age peers	0.437	0.245	-0.043	0.917	0.074
older age peers	Janos & Robinson, 1985	older age peers	0.574	0.311	-0.036	1.184	0.065
older age peers			0.052	0.083	-0.111	0.215	0.533
same age peers	Swiatek & Benbow, 1991	same age peers	-0.528	0.212	-0.944	-0.112	0.013
same age peers	Plucker, & Taylor, 1998	same age peers	0.041	0.105	-0.165	0.247	0.696
same age peers	Houston, 1999	same age peers	0.083	0.315	-0.534	0.699	0.793
same age peers	Chilton, 2001	same age peers	0.105	0.184	-0.256	0.466	0.570
same age peers	Cornell, Callahan, & Loyd, 1991	same age peers	0.110	0.290	-0.458	0.678	0.704
same age peers	Washington, 1999	same age peers	0.160	0.148	-0.131	0.451	0.281
same age peers	Lupkowski, Whitmore, & Ramsay, 1992	same age peers	0.163	0.071	0.024	0.302	0.021
same age peers	Stamps, 2004	same age peers	0.168	0.239	-0.300	0.636	0.482
same age peers	Jin & Moon, 2006	same age peers	0.664	0.130	0.408	0.920	0.000
same age peers			0.141	0.079	-0.013	0.295	0.072
Overall			0.071	0.060	-0.046	0.189	0.235

Hedges's g and 95% CI



under randomeffects model

Table 15. Summary of Combined Effect Sizes for Social-Emotional Development

Outcome group		No. of studies	Combined ES (Hedges's *g*)	Standard error(SE)	95% Confidence intervals		Test of null (2-tail) p-value
					Lower limit	Upper limit	
Social-emotional overall		22	.076	.051	-.025	.176	.139
Social-emotional	with same age peers	9	.141	.079	-.013	.295	.072
	with older age peers	10	.052	.083	-.111	.215	.533
	with mixed-age peers	3	-.036	.124	-.280	.208	.773

*$p < 0.05$ **$p < 0.01$ ***$p < 0.001$

5.2.2 Heterogeneity Analysis

Table 16 provides a summary of the heterogeneity analysis results for social-emotional development. As Table 16 indicates, there were two groups of effect sizes exhibiting high degrees of heterogeneity. These two groups were Social-Emotional Overall (g= .076, Q=43.515, df = 21, I^2=51.741, $p < .01$) and Social-Emotional with Same Age Peers (g = -.036, Q=26.928, df = 8, I^2=70.292, $p < .01$). Based on the heterogeneity analysis results above, further testing for moderators was performed on the group of Social-Emotional Overall, because this group consisted of more than 10 studies (Higgins & Green, 2006).

Table 16. Summary of Heterogeneity Analysis for Social-Emotional Development

Outcome group		No. of studies	Combined ES	Q-value	*df* (Q)	Heterogeneity *p*-value	I^2
Social-emotional overall		22	.076	43.515	21	.003**	51.741
Social-emotional	with same age peers	9	.141	26.928	8	.001**	70.292
	with older age peers	10	.052	9.981	9	.352	9.825
	with mixed-age peers	3	-.036	1.458	2	.482	.000

*p < 0.05 **p < 0.01 ***p < 0.001

5.2.3 Testing for Moderators

For the same reasons as described in the academic achievement results part in this chapter, a total of 25 variables were tested for moderators for the social-emotional development effects (see Table 17). As Table 17 indicates, effects were heterogeneous within most of the tested variables. Two variables were significantly associated with the total variance. They were acceleration form (*g* =.128, 95% CI -.044 to .300, Q_b=13. 653, *df* = 6, $p < .05$) and study time (*g* =.097, 95% CI -.099 to .293, Q_b=7. 251, *df* = 2, $p < .05$). Thus, these two variables were identified as moderators. None of the moderators of the effects of social-emotional development were the same as those of the effects of achievement as discussed in the academic achievement results.

Table 17 also shows the magnitude and the dispersion of the effect for each coding category within each tested variable. For the variable of acceleration forms, ‘early graduation’ (*g* .664, 95% CI .338 to .990, 1 study) exhibited the highest effect, followed by ‘curriculum compacting’ (*g* .168, 95% CI -.342 to .678, 1 study), ‘mentoring’ (*g* .083, 95% CI -.567 to .732, 1 study) and ‘grade-skipping’ (*g* .041, 95% CI -.248 to .330, 1 study). It is relevant to note that there was only one study contributing to each of these three coding categories. Furthermore, the results indicate that ‘early kindergarten, school or college entrance’ was the most common acceleration practice that had been researched and it showed an effect close to zero (*g* .075, 95% CI -.049 to .199, 12 studies). In addition, ‘subject-matter acceleration’ was another acceleration intervention that had been comparatively frequently studied and it also showed an effect close to zero (*g* .003, 95% CI -.183 to .188, 4 studies). The rests of two studies that fell into the category of ‘multiple forms’ also exhibited an effect close to zero (*g* .006, 95% CI -.199 to .211, 2 studies).

For the variable of study time, the results indicated that ‘studies conducted within one year after acceleration’ (*g* .260, 95% CI .105 to .415, 8 studies) had a positive effect, whereas ‘longitudinal studies’ showed an effect close to zero (*g* -.004, 95% CI -.123 to .151, 11 studies). It should also be noted that ‘studies conducted within one year after acceleration' showed statistically significant positive effects because the 95% CI of the effect size (*g* = .260) did not include zero. Three studies that fell into the category of 'both' showed a slightly positive effect (*g* .025, 95% CI -.225 to .276, 3 studies).

In addition, to determine if there was any difference between subject-based acceleration and grade-based acceleration in terms of the social-emotional development effects, a significance test was performed on the two related sub-groups. As Table 17 shows, there was one study examining both subject-based and grade-based acceleration, and no study investigating an acceleration intervention that could be categorized into either subject-based or grade-based acceleration. Thus, a total of 21 studies that either examined subject-based or grade-based acceleration remained in the test. No statistically significant differences between these two groups were seen for the effect.

Table 17. Results of Testing Categorical Variables on Social-Emotional Development

Variable	No. of studies	*g*	Lower limit	Upper limit	$Q_w(df)$	I^2	$Q_b(df)$	*p*
Study focus	**22**	**0.082**	**-.021**	**.184**			**.012(1)**	**.911**
1.Academic achievement								
2.Social-emotional development	11	.076	-.072	.223	32.835(10)	69.545		
3.Both	11	.087	-.055	.230	8.786(10)	.000		
Identification	**22**	**.080**	**-.030**	**.190**			**.590(4)**	**.964**
1.Achievement scores	5	.118	.-.112	-.348	3.279(4)	.000		
2.IQ								
3.Age								
4.Teacher/parent identification								
5.Self-reported	2	.046	-.298	.391	.081(1)	.000		
6.Multiple ways	13	.081	-.058	.219	37.345(12)	67.867		
7.Enrolled in gifted/enrichment program/class	1	-.171	-.906	.564	.000(0)	.000		
8.Classified as G/T in regional, such as states, district, schools, etc	1	.083	-.634	.799	.000(0)	.000		
9.Qualified for certain class, school, college, or grade level								
Acceleration category	**22**	**.081**	**-.022**	**.185**			**.871(2)**	**.647**
1. Grade-based acceleration	15	.115	-.010	.239	32.181(14)	56.497		
2. Subject-based acceleration	6	.019	-.183	.222	8.286(5)	39.660		
3. Both 1 and 2	1	-.023	-.442	.396	.000(0)	.000		

Variable	No. of studies	*g*	Lower limit	Upper limit	$Q_w(df)$	I^2	$Q_b(df)$	*p*
4. Non-specific								
Acceleration forms*	**22**	**.128**	**-.044**	**.300**			**13.653(6)**	**.034***
1. Early kindergarten, school or college entrance	12	.075	-.049	.199	12.664(11)	13.140		
2. Grade skipping	1	.041	-.248	.330	.000(0)	.000		
3. Advanced placement/Dual Credits/ International Baccalaureate								
4. Subject-matter acceleration (e.g., math and reading)	4	.003	-.183	.188	7.908(3)	62.062		
5. Curriculum compacting	1	.168	-.342	.678	.000(0)	.000		
6. Multiple forms	2	.006	-.199	.211	.088(1)	.000		
7. Early graduation	1	.664	.338	.990	.000(0)	.000		
8. Mentoring	1	.083	-.567	.732	.000(0)	.000		
Acceleration duration	**22**	**.080**	**-.029**	**.190**			**.510(3)**	**.917**
1.Under 14 weeks	1	.174	-.308	.656	.000(0)	.000		
2.14 to 28 weeks	2	.165	-.161	.490	.000(1)	.000		
3.Over 28 weeks	18	.061	-.063	.184	40.780(17)	58.313		
4.Non-specific	1	.105	-.408	.617	.000(0)	.000		

Variable	No. of studies	*g*	Lower limit	Upper limit	$Q_w(df)$	I^2	$Q_b(df)$	*p*
Grade level at acceleration	**22**	**.084**	**-.067**	**.235**			**8.376(7)**	**.301**
1. Elementary	3	.008	-.275	.290	2.020(2)	.973		
2. Middle school	1	.083	-.604	.769	.000(0)	.000		
3. High school	3	.358	.098	.618	8.402(2)	76.195		
4. Post secondary	8	.057	-.121	.235	7.756(7)	9.746		
5. Both 1 and 2	2	.004	-.256	.264	.088(1)	.000		
6. Both 2 and 3	2	-.146	-.452	.159	5.920(1)	83.108		
7. Both 3 and 4								
8. Kindergarten	2	.272	-.074	.618	.804(1)	.000		
9. varying (include multiple levels)	1	.041	-.323	.405	.000(0)	.000		
Study time*	**22**	**.097**	**-.099**	**.293**			**7.251(2)**	**.027***
1. Study conducted within 1 year after acceleration	8	.260	.105	.415	15.335(7)	54.353		
2. Longitudinal studies	11	-.004	-.123	.115	11.308(10)	11.568		
3. Both	3	.025	-.225	.276	3.308(2)	39.539		
School type	**22**	**.083**	**-.026**	**.201**			**1.212(4)**	**.876**
1. Rural	1	.168	-.420	.756	.000(0)	.000		
2. Suburban	2	-.003	-.340	.334	2.351(1)	57.467		
3. Urban	3	.235	-.095	.564	1.054(2)	.000		

Variable	No. of studies	*g*	Lower limit	Upper limit	$Q_w(df)$	I^2	$Q_b(df)$	*p*
4. Not applicable	15	.066	-.064	.196	37.233(14)	62.399		
5. National	1	.063	-.361	.487	.000(0)	.000		
SES	**22**	**.080**	**-.030**	**.190**			**.255(4)**	**.993**
1. Low	1	.083	-.636	.801	.000(0)	.000		
2. Medium	2	.151	-.185	.487	2.546(1)	60.720		
3. High	1	.110	-.567	.787	.000(0)	.000		
4. Non-specific	15	.076	-.056	.207	36.665(14)	61.186		
5. Low to medium	3	.039	-.259	.337	2.693(2)	25.730		
6. Medium to high								
Ethnicity	**22**	**.081**	**-.027**	**.188**			**.111(3)**	**.991**
1. Non-minority students dominant (over 60%)	4	.116	-.146	.377	4.907(3)	38.861		
2. Minority students dominant (over 60%)	1	.083	-.628	.793	.000(0)	.000		
3. Approximately equivalent	1	.041	-.367	.449	.000(0)	.000		
4. Non-specific	16	.076	-.049	.201	37.241(15)	59.722		
Gender	**22**	**.081**	**-.026**	**.188**			**.235(3)**	**.972**
1. Male predominant (over 60%)	5	.117	-.079	.313	27.924(4)	85.675		
2. Female predominant (over 60%)	6	.074	-.164	.313	5.461(5)	8.435		
3. Approximately equivalent	6	.049	-.149	.246	.773(5)	.000		

Variable	No. of studies	*g*	Lower limit	Upper limit	$Q_w(df)$	I^2	$Q_b(df)$	*p*
4. Non-specific	5	.080	-.157	.317	5.771(4)	30.682		
Study retrieval source	**22**	**.058**	**-.106**	**.221**			**2.234(2)**	**.327**
1.Journal article	15	.103	-.010	.216	33.430(14)	58.122		
2.Doctoral dissertation	6	.072	-.151	.295	5.413(5)	7.626		
3.Unpublished study obtained	1	-.306	-.832	.220	.000(0)	.000		
Sample assignment	**22**	**.082**	**-.019**	**.183**			**.093(1)**	**.761**
1. Randomized	1	.168	-.395	.731	.000(0)	.000		
2. Non-randomized	21	.079	-.023	.182	42.365(20)	52.791		
3. Mixed								
Comparison group construction	**22**	**.068**	**-.060**	**.197**			**1.676(4)**	**.795**
1. Matching on IQ, achievement scores, age, or grade-level.	13	.122	-.015	.258	35.875(12)	66.551		
2. Ex post facto design	2	-.154	-.573	.264	1.021(1)	2.094		
3. Randomization from initial pool of eligible subjects	2	.037	-.303	.377	.032(1)	.000		
4. Randomization from clustered sample groups	2	.035	-.421	.491	.705(1)	.000		
5. Self-reported information	3	.061	-.207	.329	.058(2)	.000		

Variable	No. of studies	*g*	Lower limit	Upper limit	$Q_w(df)$	I^2	$Q_b(df)$	*p*
6. The rest of eligible subjects after treated groups have been assigned.								
Subject group numbers	**22**	**.082**	**-.021**	**.184**			**.011(1)**	**.915**
1. Two groups	12	.086	-.046	.218	36.803(11)	70.111		
2. More than two groups	12	.075	-.087	.236	4.924(9)	.000		
Test design	**22**	**.081**	**-.023**	**.186**			**.196(1)**	**.658**
1.Posttest only	20	.072	-.039	.184	41.315(19)	54.012		
2. Pretest-posttest	2	.143	-.148	.434	.087(1)	.000		
3. Multiple times tests								
Research design	**22**	**.082**	**-.019**	**.183**			**.093(1)**	**.761**
1. Experimental								
2. Quasi-experimental	1	.168	-.395	.731	.000(0)	.000		
3. Causal-comparative research	21	.079	-.023	.182	42.365(20)	52.791		
Sample size	**22**	**.080**	**-.028**	**.189**			**.355(3)**	**.949**
1. ≤ 100	8	.060	-.158	.278	13.646(7)	48.703		
2. 101-500	10	.107	-.046	.261	25.308(9)	64.437		
3. 501-1000	2	.011	-.290	.312	.141(1)	.000		
4. >1000	2	.085	-.218	.387	.682(1)	.000		
5. Not applicable								

Variable	No. of studies	*g*	Lower limit	Upper limit	$Q_w(df)$	I^2	$Q_b(df)$	*p*
Statistical analysis	**22**	**.081**	**-.024**	**.186**			**.013(2)**	**.993**
1. Descriptive statistics (including dichotomous outcomes)								
2. Student-level T-test, ANOVA, or MANOVA	19	.083	-.030	.197	41.515(18)	56.642		
3. Student-level ANCOVA, MANCOVA, Logistic Regression Analysis	2	.068	-.312	.449	.856(1)	.000		
4. Cluster-level analysis								
5. Hierarchical Linear Model (HLM) analysis with cluster-level assignment	1	.063	-.349	.475	.000(0)	.000		
Data source	**22**	**.080**	**-.030**	**.190**			**.355(3)**	**.949**
1. Standardized tests								
2. Non-standardized tests	3	.015	-.252	.283	2.410(2)	17.014		
3. Survey/questionnaires	15	.104	-.037	.246	37.104(14)	62.269		
4. Combing 1, or 2, or 3, or 5.								
5. Accumulative records or archival data	3	.065	-.201	.332	.681(2)	.000		
6. National databases	1	.063	-.369	.495	.000(0)	.000		

Variable	No. of studies	*g*	Lower limit	Upper limit	$Q_w(df)$	I^2	$Q_b(df)$	*p*
Reliability information	**22**	**.082**	**-.020**	**.183**			**.305(1)**	**.581**
1. Yes	18	.096	-.017	.209	40.490(17)	58.014		
2. No	4	.023	-.208	.255	.211(3)	.000		
Measurement reliability status	**22**	**.080**	**-.029**	**.190**			**.151(3)**	**.985**
1. High reliability	10	.097	-.063	.257	33.047(9)	72.766		
2. Moderate reliability	8	.051	-.135	.237	5.285(7)	.000		
3. Low reliability	1	.110	-.565	.785	.000(0)	.000		
4. Not-available	3	.089	-.186	.364	3.196(2)	37.431		
Study validity status	**22**	**.087**	**-.060**	**.235**			**2.102(1)**	**.147**
1. Low validity								
2. Moderate validity	14	.017	-.117	.151	19.167(13)	31.819		
3. High validity	8	.168	.014	.322	21.621(7)	67.623		
ES reporting	**22**	**.082**	**-.020**	**-.184**			**.406(1)**	**.524**
1. Yes	7	.042	-.117	.201	6.008(6)	.130		
2. No	15	.110	-.024	.243	35.825(14)	60.921		
Effect size extraction	**22**	**.080**	**-.030**	**.190**			**.210(2)**	**.900**
1. Calculated from descriptive statistics	20	.076	-.044	.195	41.186(19)			
2. Calculated from inferential statistics	1	.163	-.230	.556	.000(0)			

Variable	No. of studies	*g*	Lower limit	Upper limit	$Q_w(df)$	I^2	$Q_b(df)$	*p*
3. Estimated from probabilities	1	.041	-.380	.462	000(0)			
4. Computed effect sizes in original study								

$^*p < 0.05$ $^{**}p < 0.01$ $^{***}p < 0.001$

5.2.4 Cumulative Analysis

Figure 12 shows the results of the cumulative analysis of the combined effect sizes for social-emotional development sorted by year of study. As depicted in Figure 12, as the 22 studies appeared successively, none of the accumulated effects was statistically significant, although the first two accumulated effects sizes [as the first two studies, Janos & Robins (1985), Robins & Janos (1986) were involved successively] appeared to be larger than the rest of those. There was no other obvious change pattern observed in terms of the magnitudes of the accumulated effect sizes as the newer studies appeared over the time.

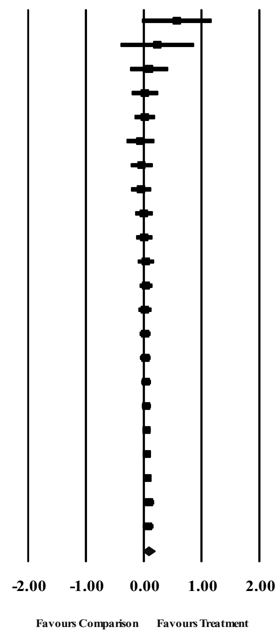
However, starting from the early 1990s through the 2000s, the standard errors associated with the accumulated effect sizes showed a decreasing pattern. Specifically, starting from the Swiatek & Benbow (1991) study to the Stamps (2004) study, 15 studies appeared. During this period, the standard errors consecutively decreased from .128 to .036, with the exceptions of two cases. It can be concluded that, overall, the standard errors decreased since the early 1990s. This suggests that research on the effects of acceleration on social-emotional development might have been conducted with higher precision since the early 1990s than before, because the researchers used larger sample sizes or more reliable measurement tools.

Figure 12. Cumulative Analysis of Social-Emotional Development Sorted by Year of Study

Cumulative Analysis of Social-Emotional Development Sorted by Year of Study

Study name	Comparison	Time point	Point	Standard error	Lower limit	Upper limit	p-Value
Janos & Robinson, 1985	older age peers	1985	0.574	0.311	-0.036	1.184	0.065
Robinson & Janos, 1986	older age peers	1986	0.234	0.329	-0.410	0.879	0.476
Brody & Benbow, 1987	mixed-age peers	1987	0.094	0.173	-0.245	0.433	0.587
Janos, Robinson, & Lunneborg, 1989	older age peers	1989	0.022	0.121	-0.216	0.260	0.854
Cornell, Callahan, & Loyd, 1991	same age peers	1991	0.020	0.095	-0.165	0.206	0.832
Swiatek & Benbow, 1991	same age peers	1991	-0.055	0.128	-0.307	0.196	0.666
Swiatek, & Benbow, 1991 (2)	older age peers	1991	-0.036	0.102	-0.235	0.163	0.723
Ingersoll, 1992	older age peers	1992	-0.046	0.092	-0.227	0.135	0.618
Lupkowski, Whitmore, & Ramsay, 1992	same age peers	1992	0.004	0.082	-0.157	0.166	0.959
Noble & Robinson, 1993	older age peers	1993	0.010	0.076	-0.138	0.159	0.891
Reynolds, 1993	older age peers	1993	0.038	0.075	-0.110	0.186	0.613
Sayley & Brookshire, 1993	older age peers	1993	0.039	0.062	-0.081	0.160	0.524
Ambruster, 1995	mixed-age peers	1995	0.020	0.060	-0.097	0.138	0.737
Plucker, & Taylor, 1998	same age peers	1998	0.026	0.052	-0.077	0.128	0.621
Houston, 1999	same age peers	1999	0.029	0.050	-0.069	0.127	0.559
Washington, 1999	same age peers	1999	0.040	0.046	-0.051	0.131	0.385
Chilton, 2001	same age peers	2001	0.045	0.043	-0.040	0.131	0.295
Ma, 2002	mixed-age peers	2002	0.050	0.039	-0.028	0.127	0.208
Gagné & Gagnier, 2004	older age peers	2004	0.057	0.037	-0.016	0.130	0.128
Stamps, 2004	same age peers	2004	0.061	0.036	-0.009	0.131	0.086
Jin & Moon, 2006	same age peers	2006	0.091	0.051	-0.010	0.191	0.076
Sayler, 2008	older age peers	2008	0.076	0.051	-0.025	0.176	0.139
			0.076	0.051	-0.025	0.176	0.139

Cumulative hedges's g (95% CI)



under randomeffects model

5.2.5 Assessment of Publication Bias

The assessment of publication bias for the meta-analysis of social-emotional development effects was conducted on all 22 studies. Among these studies, 15 (68.2%) were published journal articles, 6 (27.3%) were doctoral dissertations, and 1 (4.5%) was a current study obtained from researchers. Figure 13 displays a funnel plot for a visual inspection of the publication bias. As Figure 13 shows, the majority of the studies symmetrically appeared around the mean effect line of the funnel plot. This suggests that there was no presence of publication bias. However, the lower precision shown in the funnel plot also indicated that, just like those in the meta-analysis of academic achievement effects, small studies once again played a major role in the meta-analysis of social-emotional development effects. This leads to a further confirmation to the perception about the small sample size issues in gifted/talented education research.

Figure 14 shows a funnel plot after Duvall and Tweedie's trim and fill procedure. This contains information for further assessing the publication bias. As can be seen in Figure 14, this funnel plot looks nearly identical to the plot before the trim and fill procedure. The statistical information yielded with the trim and fill procedure also showed that the point estimate and the 95% confidence interval for the observed and adjusted effects were the same: .093 (-.025, .176). Therefore, it can be concluded that publication bias was not of concern in this case.

Figure 13. Funnel Plot of Precision by Hedges's *g* (Social-Emotional Development)

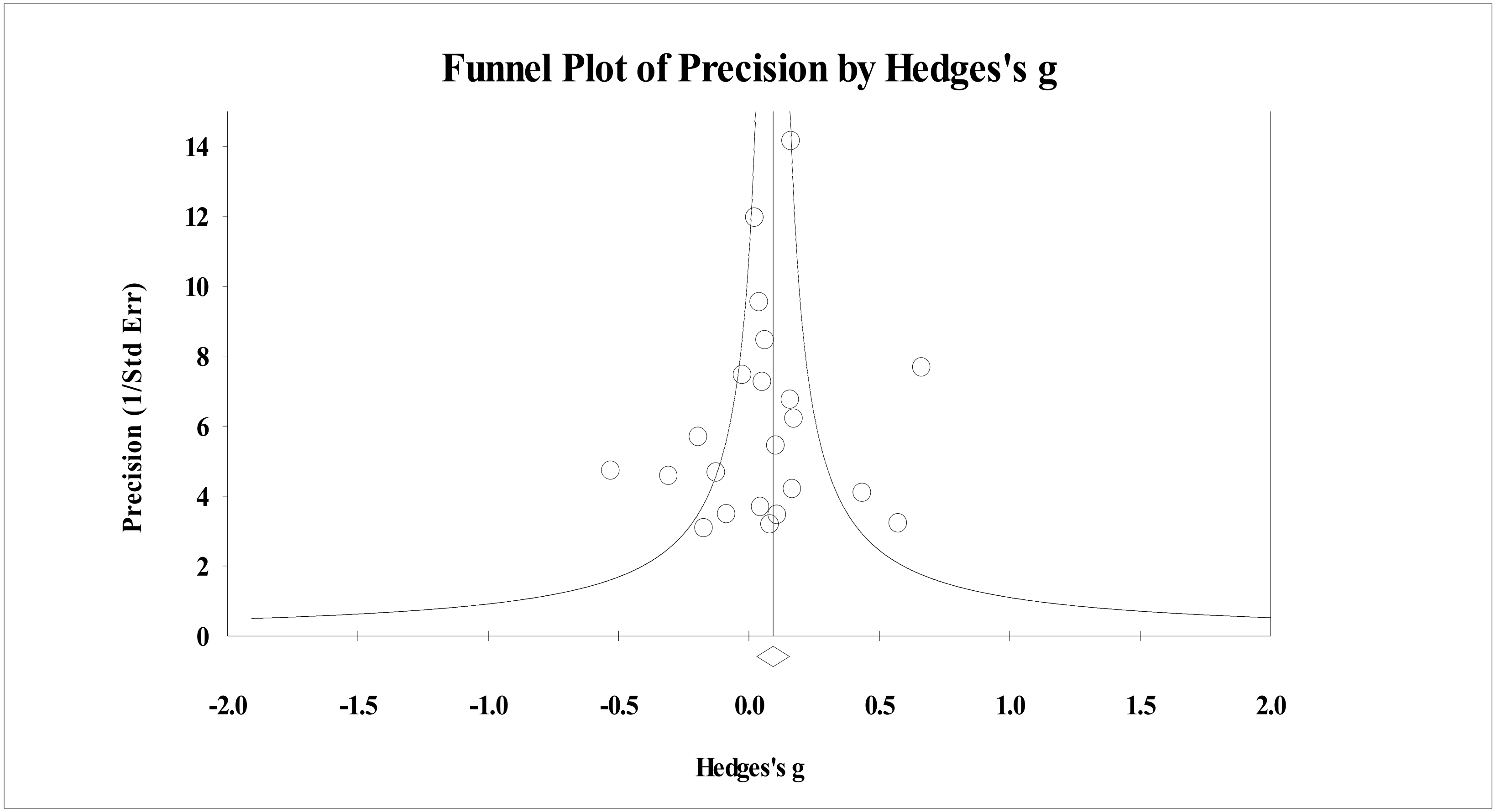
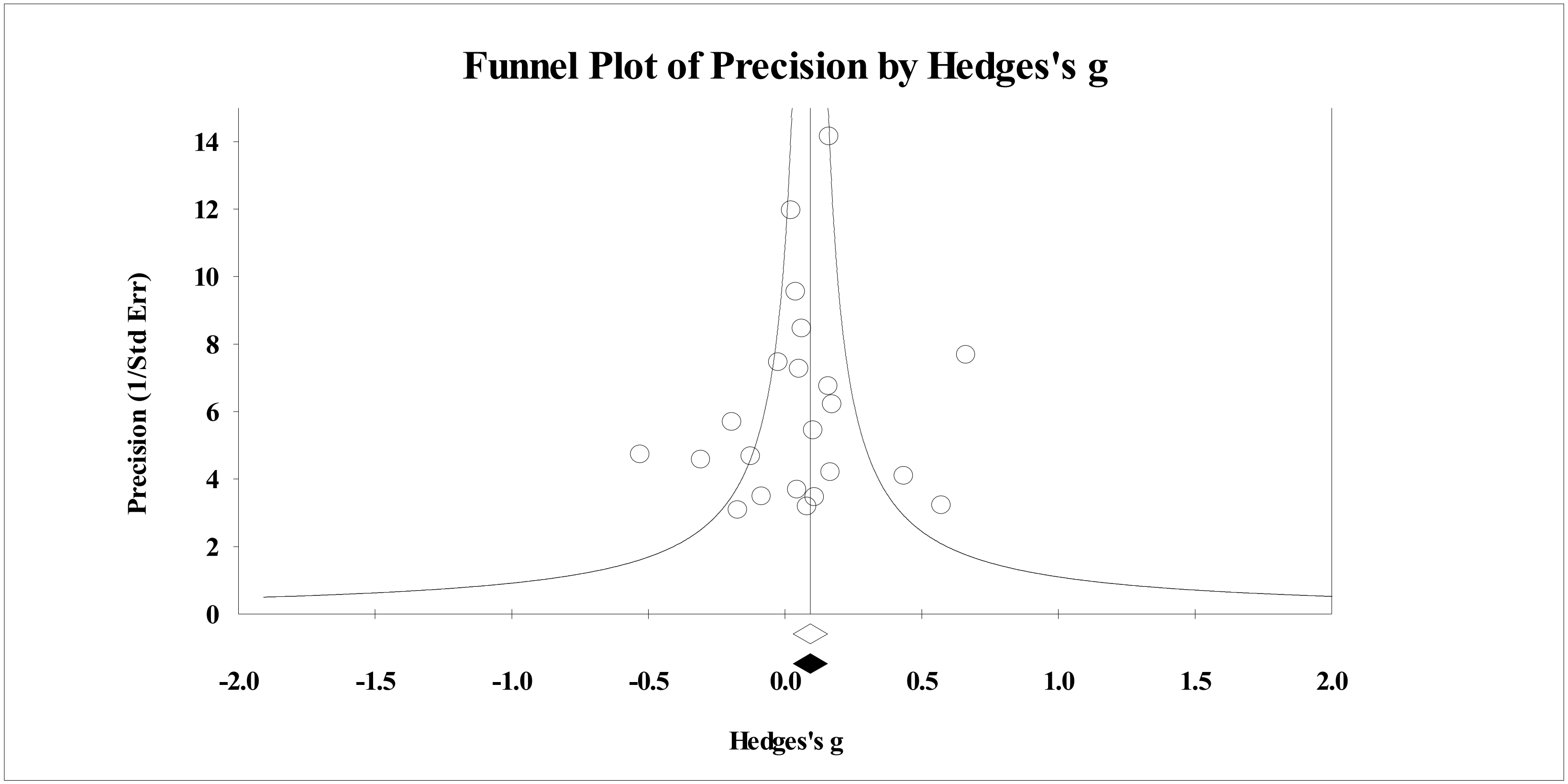


Figure 14. Funnel Plot after Trim and Fill Procedure (Social-Emotional Development)



CHAPTER 6 DISCUSSION AND CONCLUSIONS

This chapter summarizes and discusses the key findings from this meta-analysis. First, effect size interpretation guidelines are discussed. Then, each of the six research questions are addressed based on the findings from this study. Study limitations and implications for future research and educational practice are discussed. Finally, overall conclusions are drawn.

6.1 Summary and Interpretation of Findings

6.1.1 Effect Size Interpretation Guidelines

Effect sizes are informative only when they are interpreted appropriately. According to Thompson (2006), a typical method to interpret quantitative inquiry results is to evaluate some combination of the practical, statistical, or clinical significance of the results. The interpretation of the results of this meta-analysis was conducted using the following two guidelines: First, the results of the study were interpreted in terms of their practical significance, rather than their statistical or clinical significance. This is because, generally speaking, the perspective of practical significance fits the context and nature of educational research better than those of either statistical or clinical significance. An examination of previous meta-analytic studies on the effects of acceleration revealed a better understanding of the importance of interpretation perspectives. For example, in Kulik and his colleagues' earlier meta-analysis of acceleration (i.e., Kulik & Kulik, 1984), results were simply interpreted using Cohen's (1977) guidelines, in which no differentiation of interpretation perspectives was meant or implied, whereas in Kulik's (2004) meta-analysis of acceleration, results were interpreted in terms of practical significance. Interpreting the results of this meta-analysis in terms of practical significance would enable one to closely compare this study with Kulik's (2004) meta-analysis and draw appropriate conclusions.

Second, the results were compared with those of related previous meta-analytic studies on the effects of acceleration, rather than being rigidly interpreted under fixed benchmarks, such as Cohen's effect size interpretation guidelines. This is because referring to the research context and the specific study questions is commonly considered a better approach than following a fixed benchmark for effect size interpretation. For example, Glass, McGaw and Smith (1981) claimed that, at least in relatively established areas, "there is no wisdom whatsoever in attempting to associate regions of the effect-size metric with descriptive adjectives such as 'small,' 'moderate,' 'large,' and the like" (p.104). Vacha-Haase and Thompson (2004) stated that effect sizes should be interpreted, first, by taking into account the specific outcome being studied, and then by "directly and explicitly comparing effects with those in related prior studies" (p. 478). Cooper (2007) also noted that it is most informative to compare the effect size values with those of closely related research.

In principle, a meta-analytic perspective was adopted in results interpretation in this meta-analysis, using what some have called "meta-analytic thinking" (cf. Cumming & Finch, 2001; Thompson, 2002). As Thompson (2006) recommended, two questions were asked when interpreting effect sizes in this study with a meta-analytic perspective: (a) was the effect size noteworthy? and (b) were the effects consistent with the related prior literature? It should be noted that in educational research, even small effect sizes may be considered noteworthy, because in many cases multiple factors contribute to an observed outcome and a single or even a few interventions may be found to have quite limited impact on the outcome (Ahadi & Diener, 1989; Berliner, 2002; Strube, 1991).

Specifically, in this meta-analysis, the results were interpreted and discussed in terms of the direction and magnitude of the combined (average) effects, the width and distributions of their confidence intervals (CIs), and heterogeneity of effects, along with any concerns about observed publication bias and study validity. In accordance with the purpose of this meta-analysis, the interpretation of the results was carried out by addressing the six research questions.

6.1.2 Research Question 1

How does acceleration affect high ability learners' academic achievement?

Findings from this meta-analysis suggested that, overall, acceleration does improve high-ability learners' academic achievement. The findings revealed that acceleration had a positive influence on high-ability learners' academic achievement (*g* .180, 95% CI -.072 to .431), although the heterogeneity analysis revealed that the effect sizes within the group of Achievement Overall exhibited a high degree of heterogeneity. The 28 effect sizes (representing 28 studies) that contributed to the combined effects ranged from – 2. 493 to 1.809, with .294 as the median effect size. Five effect sizes were negative effect sizes and the remaining 23 were positive. Effect sizes most frequently occurred in the .301 to .353 range, with 7 effect sizes in this category. The related cumulative analysis suggested that the accumulated effect sizes showed a decreasing pattern since the 2000's. An examination of the methodological features of the included studies indicated that studies were conducted with more rigorous methods since the 2000's. Moreover, the assessment of publication bias suggested that publication bias was present and it was possible that the combined effect size may be overestimated. However, further investigation of publication bias suggested the publication bias was not caused by the small sample size effects alone in this case. Overall, findings from this study are consistent with those from previous meta-analytic studies, such as Kulik and Kulik (1984), Rogers (1991), and Kulik (2004), in which a statement was typically made that acceleration had a significant positive impact on high-ability learners.

When the academic achievement effects were assessed in terms of outcome levels, positive effects were found for both the P-12 Achievement (*g* .147, 95% CI -.174 to .467, 21 studies) and the Post-secondary Achievement (*g* .313, 95% CI -.262 to .889, 7 studies). Because the perspective of investigating the acceleration effects by outcome levels was first explored in this meta-analysis, no comparison with previous meta-analytic studies was possible.

The investigation of comparison groups showed the greatest effects for the three comparisons with same age peers (i.e., g = .396, g =.347, and g =.498, respectively). This suggests that the effects of acceleration may be more discernable when accelerated high-ability learners are compared with their non-accelerated same age peers. This finding is also consistent with those from some previous meta-analytic studies. For example, in both Kulik and Kulik (1984) and Kulik (2004) meta-analyses, a median effect size of .80 was found when same age comparison groups were used, whereas an effect size of -.04 was yielded when older age comparison groups were used.

In terms of the sub-group of ‘with older age peers’, there is discrepancy between the findings from this study and those from prior meta-analytic studies. Specifically, when the Academic Achievement Overall, P-12 Achievement and Post-secondary Achievement were sorted by comparison groups, the sub-groups of ‘with older age peers’ exhibited positive effects in all three analyses (i.e., g = .224, g = .334, g = .255, respectively). However, in two previous meta-analytic studies (i.e., Kulik & Kulik, 1984; Kulik, 2004), negative effects were found when accelerated gifted students were compared with older age peers.

It should be emphasized that inferences derived from the subgroup analyses should be viewed with caution and need to be replicated by future research with more studies to be considered robust. In addition, when only a limited number of studies were involved in the sub-group analysis, studies with extreme effect sizes, such as the Hsu (2003) study, may have had a strong influence on the results. This is particularly troubling in the case of Hsu (2003) because this study was categorized as low in validity (though it was high in precision).

6.1.3 Research Question 2

How does acceleration affect high ability learners’ social-emotional development?

The results from this meta-analysis suggest the effects of acceleration on high-ability learners’ social-emotional development were slightly positive, although not as positive as on academic achievement. For example, the overall combined effect size for social-emotional development was .076 (95% CI -.025 to .176, 22 studies). Again, the effect sizes associated with this combined effect exhibited a high degree of heterogeneity. However, comparing with academic achievement, the heterogeneity of effect sizes in social-emotional groups appeared to be a little lower. No extreme effect sizes existed. The 22 effect sizes that contributed to the combined effects ranged from – .528 to .664, with .058 as the median effect size. Seven effect sizes were negative and the remaining 15 were positive. Effect sizes most frequently occurred in the .024 to .083 range (6 effect sizes) and the .105 to .174 range (6 effect sizes). The assessment of publication bias suggested that publication bias was not of concern for the social-emotional findings. Furthermore, compared to prior meta-analytic studies, a more positive impression of the effects of acceleration on social-emotional development was found in this meta-analysis, perhaps because the related cumulative analysis revealed that research on social-emotional development effects of acceleration might have been conducted with higher precision (indicated by smaller standard errors) since the early 1990s than before.

When combined effects were assessed in terms of comparison groups, the effects appeared mixed. Specifically, comparisons with same age peers yielded a positive effect (*g* .141, 95% CI -.013 to .295, 9 studies), followed by those with older age peers (*g* .052, 95% CI -.111 to .215, 10 studies); whereas comparisons with mixed-age peers exhibited a slightly negative effect (*g* -.036, 95% CI -.280 to .208, 3 studies). However, the latter group had very few studies so the results may be different with future studies. Overall, the results derived from this meta-analysis were slightly more positive than previous meta-analytic studies in which claims were made that the acceleration had mixed effects on social-emotional development (e.g., Kulik, 2004) or did ‘no harm’ (e.g., Kent, 1992).

6.1.4 Research Question 3

Which conclusions from previous meta-analytic studies regarding the effects of acceleration supported?

In brief, a conclusion from the four prior meta-analytic studies (summarized in Table 3 of Chapter 2) is that acceleration has significant positive impact on high-ability learners' academic achievement, and does 'no harm' to their social-emotional development. Overall, the current study supports that conclusion. However, the greater precision of the studies in this meta-analysis and the slightly positive overall finding suggested that the social emotional effects of acceleration may be more positive than previously thought. At the very least, the results of this study support the notion that acceleration is not harmful to social-emotional development.

6.1.5 Research Question 4

What differences exist between subject-based acceleration and grade-based acceleration in terms of their effects on high-ability learners?

In terms of academic achievement effects, answers to this question varied. First, no statistically significant difference between the effects of subject-based acceleration and grade-based acceleration was detected for P-12 students as a whole ($Q_b = .292$, $df = 1$, $p > .05$). However, comparisons with same-age peers yielded a statistically significant difference ($Q_b = 6.786$, $df = 1$, $p < .01$). Somewhat surprisingly, the results indicted subject-based acceleration had much stronger positive effects on achievement (*g* 1.809, 95% CI .671 to 2.947, 11 studies) than did grade-based acceleration (*g*. 269, 95% CI -.562 to 2.429, 1 study), but this may be because there was only one study in the grade-based acceleration category.

It must be emphasized that the interpretations of these findings should be tentative rather than conclusive, considering the limited number of studies involved in the significance tests, the very small number of studies in some of the subgroups, and the high degrees of heterogeneity within the tested groups. Moreover, one should bear in mind that the coding of 'subject-based' versus 'grade-based' acceleration was not free from subjectivity as no clear cut line between the two had been drawn in the existing acceleration research literature. For example, it was hard to determine if mentoring should be coded as subject-based or grade-based acceleration. As a result, studies investigating the effects of mentoring programs were coded as 'no-specific' and were excluded from the analysis of testing the differences between the effects of subject-based and grade-based acceleration. Finally, because no previous meta-analytic studies provided any information comparing types of acceleration, no comparison was made between this meta-analysis and the previous ones.

In terms of the social-emotional effects, no statistically significant difference between subject-based acceleration and grade-based acceleration was seen for the combined effects (Qb = 4.042, *df* = 1, p > .05). Again, this result should be interpreted with caution because there was only one study in the grade-based acceleration group, and no comparison could be made between this meta-analysis and the previous ones.

6.1.6 Research Question 5

What are the typical effects of the most common acceleration forms, such as grade-skipping, early kindergarten/school/college entrance, advanced placement, subject-matter acceleration (e.g., math and reading), and curriculum compacting on high-ability learners?

The typical effects on academic achievement related to this question came from the analyses of testing for moderators in two groups: P-12 students and comparisons with same age peers (see Tables 12 & 13). For P-12 students, curriculum compacting exhibited the strongest positive effect, though the results were based on only 1 study (*g* .353, 95% CI -1.413 to 2.119, 1 study). Early kindergarten/school/college entrance (*g* .254, 95% CI -.544 to 1.053, 5 studies) and advanced placement (*g* .206, 95% CI -1.544 to 1.956, 1 study) also had positive effects, though, again, the AP results were based on only one study. No conclusions about the effects of grade-skipping can be drawn here because no studies of grade skipping were included at the P-12 level. Somewhat surpisingly, subject-matter acceleration appeared to have a negative influence on accelerants (*g* -.005, 95% CI -.639 to .629, 8 studies). It should be noted that the result about the effects of subject-matter acceleration may not be valid because two studies that had extreme effect size values (i.e., Callahan & Smith, 1990; Hsu, 2003) happened to fall into this category.

In comparisons with same age peers, positive effects were found for subject-matter acceleration (*g* .356, 95% CI .019 to .692, 7 studies), curriculum compacting (*g* .353, 95% CI -.464 to 1.170, 1 study), and advanced placement (*g* .151, 95% CI -.395 to .698, 2 studies). The most reliable effect here was for subject-matter acceleration, which was based on 7 studies. This analysis did not include the studies with extreme effect sizes mentioned above. An extremely large positive effect was found for early kindergarten/school/college entrance (*g* 1.809, 95% CI .589 to 3.029, 1 study), but this result was based on only one study that had a large effect size (i.e., Janos, 1987). Again, the effects of grade skipping were not investigated.

In terms of social-emotional effects, the following points can be made tentatively: (a) early kindergarten/ school/college entrance exhibited a slightly positive effect (*g* .075, 95% CI -.049 to .199, 12 studies) on a wide variety of social/emotional outcomes (e.g., self-concept, self-esteem, locus of control, and peer perceptions) and has been the primary focus of research on the social/emotional effects of acceleration, (b) curriculum compacting had a positive effect (*g* .168, 95% CI -.342 to .678, 1 study), as did grade-skipping (*g* .041, 95% CI -.248 to .330, 1 study), but these conclusions are more tentative because they are based on single studies. Subject-matter acceleration had little effect on social/emotional development across four studies (*g* .003, 95% CI -.183 to .188, 4 studies). Very few studies contributed to the combined effects for each of most of these forms of acceleration and the social/emotional outcomes measured varied tremendously across studies. The effects of advanced placement on social-emotional development were not investigated. With the exception of ‘early kindergarten/ school/college entrance’, no comparison could be made between this meta-analysis and the previous ones. Kent's (1992) meta-analysis yielded a slightly higher effect size (*d* =.14) for early kindergarten entrants. A conservative conclusion is that this meta-analysis yielded a similar result to the prior one (i.e., Kent, 1992) and provides somewhat more information about the effects of different types of acceleration on social/emotional development. However, the results of this study should be considered quite tentative, rather than conclusive, given the fact that these results were derived from a very limited number of studies. In other words, the research base on acceleration was not sufficient to enable a comprehensive and robust investigation of this research question.

6.1.7 Research Question 6

What moderators are significantly associated with the effects of acceleration on high-ability learners?

In terms of academic achievement effects, four variables were identified as moderators of the effects on P-12 students. They were acceleration duration, comparison group construction, statistical analysis, and study validity status. It must be noted that the interpretation of the results of testing for moderators is always observational and tentative. Many factors may account for the results, for example, the number of studies involved in each coding category, the methodological features of the primary studies, the features of the intervention itself, or the analysis model used to combine the effects (e.g., random-effect or fixed-effect model), etc. As such, caution is needed to reach any justifiable conclusions based on the results of testing for moderators.

Among these four moderators, three were coded as methodological feature variables (comparison group construction, statistical analysis, and study validity), and acceleration duration was coded as a general feature. This suggests that it would be inappropriate to ignore the primary studies' methodological issues when assessing the effects of acceleration on P-12 level high-ability learners. Based on the results above, it would be sensible to conclude that comparison group construction, statistical analysis procedures, as well as the study validity are all significantly associated with the effects of acceleration. Therefore, careful consideration of these three factors is warranted for the designs of future primary studies and meta-analyses.

When testing for moderators was conducted on studies comparing accelerants to non-accelerated same age peers, nine moderators were found. They were: identification, acceleration category, acceleration duration, grade level at acceleration, SES, ethnicity, study retrieval source, statistical analysis, effect size reporting. Among these nine moderators, the first 6 were coded as general feature variables, while study retrieval source, statistical analysis and effect size reporting were variables representing methodological features of studies. This suggests that when comparing high-ability learners with their same age peers, the effects of acceleration would be affected not only by the intervention features but also by the primary study's methodological features. Because the analysis included only 13 studies and some variables included only 1 study in some categories, only a few identified moderators can be considered to have practical implications. These moderators include acceleration duration, study retrieval source and effect size reporting. It is interesting to note that acceleration duration and statistical analysis were identified as moderators in both analyses. The importance of these two factors to future acceleration research and practices can be reflected from this information.

With regard to social-emotional development, acceleration form, and study time were significantly associated with the effects and thus identified as moderators. It can be seen that these two moderators do not overlap with any of the moderators of academic achievement effect. This may suggest that it is possible that a variable that moderates academic achievement may not moderate social-emotional development, and vice versa.

6. 2 Study Limitations

The interpretation and application of the results from this meta-analysis needs to be undertaken with appropriate understanding of the limitations of this study. Aside from some general limitations associated with the methods of meta-analysis (as discussed in chapter 1), this study is subject to the following four particular limitations.

First, no strong evidence of reliability in study coding was provided from this study. Although the obtained inter-rater reliability is close to satisfactory (r =.75), the confidence level of this information is quite restricted, because only a very limited number of studies (2 out of 38) were double-coded to obtain the inter-rater reliability. Ideally, meta-analyses should be conducted with sufficient team support so that strong evidence of coding reliability can be achieved. As a doctoral dissertation, this meta-analysis was conducted primarily with the author's individual efforts and without sufficient team support, due to the sophisticated coding training process and the large amount of time needed to finish coding many studies.

A few additional procedures were carried out to increase the coding reliability. For example, the coding protocol was reviewed by an expert in gifted and talented studies and a meta-analysis expert and refined several times based on their feedback. Further, pilot coding was performed on a set of studies and further revision was conducted based on the pilot experiences. Thus, the coding protocol itself can be considered as a well-structured coding instrument. Moreover, during the coding process, a second check was performed when the coded text information was further coded into numerical data. No minor errors were found. These efforts increase the study coding reliability.

Second, a sensitivity analysis in which results derived from different analysis models can be compared to determine if the findings are robust was not undertaken in this meta-analysis. As described in the methods section, it is believed that a random-effects model would fit better than a fixed-effects model for this meta-analysis, as a result no analysis was conducted with the fixed-effects or mixed-effects models and no comparison was made thereafter. It should be noted, however, Duvall and Tweedie's trim and fill procedures, which were utilized, can be considered an alternative sensitivity analysis.

Third, this meta-analysis is subject to problematic definitions of acceleration practices. This limitation is closely related to the existing situation in the acceleration research literature. One example would be that researchers usually provided no clear definitions of the acceleration interventions they investigated, nor did they give a detailed description about the acceleration practices in their study. This posed some difficulties for study coding in this meta-analysis. Furthermore, researchers usually claimed their study effects were due to the acceleration intervention that the subjects experienced when the studies were conducted, while ignoring what happened previously. For instance, researchers who chose early college entrants as their study subjects would typically claim that their study was designed to investigate the effects of early college entrance on high-ability learners, regardless of the fact that some of their subjects were able to enter college early because they had at least one other acceleration experience earlier in their educational career, such as grade-skipping, subject-matter acceleration, or advanced-placement, etc. It seems like that the observed effect in most studies of postsecondary students might involve an accumulated effect of multiple acceleration experiences, but this was not investigated in any of the studies.

In this study, coding decisions were based upon the original claims in the primary studies, although efforts were made to inspect the available descriptions about the acceleration intervention in the study in order to verify the original claims. Therefore, findings from this meta-analysis may only serve as a rough picture about the effects of acceleration, especially when it comes to the effects of any specific acceleration forms or categories. Solutions to this limitation require researchers to provide more precise and complete descriptions of the acceleration interventions they study.

Finally, the findings from this meta-analysis do not represent all the potential information that the included studies could provide. In other words, more findings could be drawn from the coded data with additional effort. As previously described in chapter 4, a total of 274 preliminary effect sizes were extracted from the 38 studies included, serving as a rich source of data. In this study, separate meta-analyses were conducted on the effects of academic achievement and social-emotional development. As a result, findings were derived from this study in terms of these two outcome types. Within each of the outcome types, however, information regarding specific outcome variables was available. For example, for academic achievement, further meta-analyses could be conducted to obtain the combined effect in terms of some most common outcome variables, such as standardized achievement test results, college GPA, or education background. With regard to social-emotional development, there was even more variation in the measurement of outcomes, so meta-analyses could be conducted on some comparatively common sub-constructs or variables, such as social relationships, participation in extracurricular activities, mental maturity/adjustment, locus of control, life satisfaction, self-concept, self-acceptance, self-esteem, self-confidence, and educational/vocational plans. Kulik (2004) is a good example of this type of meta-analysis. These findings, if obtained, would be very informative to future acceleration research and practices.

6.3 Implications for Future Research

One implication of this study for future research is that researchers in gifted and talented education need to do a better job of specifying important demographic information. For example, it was found that among the total of 38 primary studies included in this meta-analysis, in more than two-thirds of the studies (71.1%), researchers presented no specific information about the ethnicity of the treatment samples; in close to two-thirds of the studies (68.4%), researchers provided no specific information about the SES of the study subjects; in more than half of the studies (63.2%), researchers reported no specific information about school type; in about one-third of the studies (28.9%) researchers provided no information about the gender ratio of the treatment groups. As a result, questions still remain as to the effects of acceleration on high-ability learners in terms of school context, student SES, ethnicity, and gender distribution.

This may be considered as a possible explanation for one of the conclusions in the Templeton Report on Acceleration (Colangelo et al., 2004), which stated that throughout the nation, the effects of acceleration remained vague to educational administrators, teachers and parents, although considerable evidence regarding the benefits of acceleration had been documented in the literature. Further, this also suggests that acceleration research, like educational research overall, needs to be turned in a direction in which research is better linked to the the education system practically. As Burkhardt and Schoenfeld (2003) argued, educational researchers should structure future research in a way that more useful information is accessible to practitioners and policymakers and so that evidence-based decisions can be made to improve practice.

A new research area suggested by this meta-analysis is the investigation of the impact of acceleration on high-ability learners during their transition from high school to college. Among the 38 included primary studies from 1984 to 2008, no study investigated the effects of acceleration on high-ability learners in both high school and college. A message from this information is that researchers should initiate research efforts to explore this important transition. One can expect that these research efforts would be in line with the current increasing attention on high school to college transitions in the United States. Accompanying with this high school to college pathway trend, more options are open for high-ability learners. For example, they can participate in an early college high school, take AP classes, choose to enter college early, skip high school completely, or take dual enrollment classes in their high school. As a result, more and more high-ability learners will be experiencing a nontraditional pathway between high school and college. They may have different experiences in high school, during the transition, or even after the transition. New research on the effects of acceleration during these transition periods is needed.

The criterion of requiring appropriate comparison groups played a significant role in determining the number of studies included in this meta-analysis. As previous described in chapter 4 (see Table 7), a number of studies which met all the other criteria, but failed to meet this one, were excluded. A comparison of this meta-analysis with another ongoing best-evidence synthesis of research on academic acceleration conducted by Rogers, Young and Lonergan (2008) revealed that a set of studies that were included in the latter were not qualified for being included in this meta-analysis, simply because there were no appropriate comparison groups used and statistical controls were not applied. It would be sensible to conclude that researchers should make sure that future research is conducted with appropriate comparison groups or with necessary statistical control procedures.

6.4 Implications for Practice

Students and parents may find this meta-analysis helpful for their decision-making about acceleration. As the end users of acceleration interventions, students' and parents' biggest concerns about acceleration related to the short-term and long-term influence on students' academic achievement and social-emotional development. Often times, through their personal experience, students and parents can understand and appreciate the short-term benefits of acceleration. However, they remain unsure or unconvinced about the long-term impact of their acceleration decisions. It is always the case that the evidence of the long-term effects of any educational interventions would have to rely on comprehensive practices across different contexts, various subjects, and a long period of practices. In other words, the evidence eventually needs to be based on longitudinal/retrospective research studies. Longitudinal/retrospective research studies of the effects of acceleration were scarce when the prior meta-analytic studies were conducted. In the past two decades, more and more longitudinal studies were conducted as acceleration research advance. As has been described previously, half of the studies (19 out of 38) included in this meta-analysis were longitudinal/retrospective studies. In other words, the findings of this meta-analysis are supported by a number of longitudinal studies which provided important and valuable information about the long-term effects of acceleration. Armed with this synthesized new information, students and parents will become more confident when they make decisions about acceleration.

A message from this meta-analysis for educators is that acceleration can be a valuable bridge to help them build and maintain optimal cooperation between P-12 and post-secondary educational systems. University-based acceleration programs played a very important role in implementing acceleration practices in this study. Study results showed that, overall, students viewed their experiences in the university-based acceleration programs, such as early college entrance, quite positively, and they believed that they benefited greatly in academic and social-emotional development. These positive outcomes will motivate educators in P-12 systems to devote more efforts to acceleration practices and to work more closely with universities to ensure appropriate acceleration opportunities during the transition to college. Similarly, the results of this study suggest that university-based acceleration programs need support and cooperation from educators in P-12 system, to recruit more qualified high-ability learners and achieve further success. One can expect that such a supply and demand relationship, based on successful acceleration programs, would benefit both P-12 and higher education greatly.

Policy-makers can also benefit from the findings of this meta-analysis. First of all, this meta-analysis supports the value of acceleration for high ability learners and provides additional evidence-based support for state and local acceleration policies. Second, the results of moderator exploration in this meta-analysis may be informative to policy-makers. For example, in this meta-analysis, acceleration duration was identified as one of the factors moderating the effects of acceleration. Specifically, the results showed that acceleration interventions that lasted more than 28 weeks appeared to have the greatest positive influence on high-ability learners. This is very important information that policy-makers need to take into account when planning acceleration programs. A program that is too short may not produce the best results for participants. Third, it can be seen from this meta-analysis that early entrance forms of acceleration, such as early entrance to kindergarten, school, or college programs exhibited success overall. This is the kind of information that policy-makers need to know for future practice. Finally, this meta-analysis implies that policy-makers should urge acceleration practitioners or other educators to keep good records of student demographic information and make this information more accessible for future researchers.

6.5 Conclusions

With the aim of updating previous meta-analytic studies and comprehensively synthesizing the current research findings regarding the effects of acceleration, this meta-analysis contributes to the field of gifted education in the following three ways. First, the findings from this meta-analysis generally confirm the positive influence of acceleration on high-ability learners, not only in terms of academic achievement, but also for social-emotional development. Second, results of this meta-analysis suggest that researchers should particularly pay attention to ensure that future research will be conducted with appropriate comparison groups and more research is needed to investigate the impact of acceleration on high-ability learners during high school to college transition stage. Third, newly synthesized information based on a number of longitudinal studies about the long-term effects of acceleration was derived from this meta-analysis. In summary, while supporting the generally positive effects of acceleration as found in previous research, this meta-analysis resulted in new information about acceleration that can benefit researchers, students, parents, educators, and policy-makers.

LIST OF REFERENCES

LIST OF REFERENCES

Ahadi, S. & Diener, E. (1989). Multiple determinants and effect size. *Journal of Personality and Social Psychology, 56*, 398-406.

Aiken, L.R. (2004). *Psychological testing and assessment*. New York: Allyn and Bacon.

Algina, J., & Keselman, H. J. (2003). Approximate confidence intervals for effect sizes. *Educational and Psychological Measurement, 63*, 537-553.

*Ambruster, R. L. (1995). *The effectiveness of an early entrance screening instrument as an accurate predictor of success in elementary grades*. Unpublished doctoral dissertation, Southern Illinois University at Edwardsville, AAT 9618636.

Arends, R. H., & Ford, P. M. (1964). Acceleration and enrichment in the junior high school: A follow-up study. Olympia: Washington Office of the State Superintendent of Public Instruction. (ERIC Document Reproduction Service No. ED 001 220)

Asher, J. W. (1986). Conducting research with meta-analysis: A new direction for gifted education. *Gifted Child Quarterly, 30*, 7-10.

Asher, J. W. (2003). Meta-analysis and gifted education. *Journal for the Education of the Gifted, 27*(1), 7-19.

Assouline, S. G., Colangelo, N.,Lupkowski-Shoplik, A., Lipscomb, J., & Forstadt, L. (2003). *The Iowa Acceleration Scale, 2nd Edition: Manual*. Scottsdale, AZ: Great Potential Press.

*Barnett, L. B., & Durden, W. G. (1993). Education patterns of academically talented youth. *Gifted Child Quarterly, 37* (4), 161-168.

Baum, S., & Owen, S. (1988). High ability/learning disabled students: How are they different? *Gifted Child Quarterly, 32*, 321-326.

Baurind, D. (1989). Rearing competent children. In W. Damon (Ed.), *Child development today and tomorrow*. San Francisco: Jossey-Bass.

Bax, L., Yu, L-M., Ikeda, N., & Moons, K. GM. (2007). A systematic comparison of software dedicated to meta-analysis of causal Studies. *BMC Medical Research Methodology,* published on September 10, 2007 as doi:10.1186/1471-2288-7-40, from http://www.biomedcentral.com/content/pdf/1471-2288-7-40.pdf

Benbow, C. P., Lubinski, D., Shea, D. L., & Eftekhari-Sanjani, H. (2000). Sex differences in mathematical reasoning ability at age 13: Their status 20 years later. *Psychological Science, 11*, 474-480.

Berliner, D. C. (2002). Educational research: The hardest science of all. *Educational Researcher, 31*(8), 18-20.

Bleske-Rechek, A., Lubinski, D., & Benbow, C. P. (2004). Meeting the educational needs of special populations: advanced placement's role in developing exceptional human capital. *Psychological Science, 15*, 217-224.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2006). *Comprehensive Meta-Analysis* (Version 2.2.027) [Computer software]. Englewood, NJ: Biostat

Borenstein, M., Hedges, L., & Rothstein, H. (in press). Introduction to meta-analysis. Chichester, U. K.: John H. Wiley & Sons.

Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2002). Comprehensive school reform and student achievement: A meta-analysis. Baltimore: CRESPAR, Johns Hopkins University. (ERIC Document Reproduction Service No. ED 472 569).

*Brody, L. E. (1985). *The effects of an intensive summer program on the SAT scores of gifted seventh graders (aptitude, acceleration)*. Unpublished doctoral dissertation, The Johns Hopkins University, Baltimore.

*Brody, L., Assouline, S. G., & Stanley, J. C. (1990). Five years of early entrants: Predicting successful achievement in college. *Gifted Child Quarterly, 34*, 138-142.

*Brody, L. E., & Benbow, C. P. (1987). Accelerative strategies: How effective are they for the gifted? *Gifted Child Quarterly, 31*(3), 105-110.

Brody, L. E., Muratori, M. C., & Stanley, J. C. (2004). In N. Colangelo, S. G. Assouline, & M. U. M. Gross (Eds.), *A nation deceived*: *How schools hold back America's brightest students, volume II* (pp.97-107). Iowa City, IA: The Connie Belin and Jacqueline N. Blank International Center for Gifted Education and Talent Development.

Brody, L. E., & Stanley, J. C. (1991). Young college students: Assessing factors that contribute to success. In W. T. Southern & E. D. Jones (Eds.), *The academic acceleration of gifted children*. (pp.102–132). New York: Teachers College Press.

Burkhardt, H. & Schoenfeld, A. H. (2003). Improving educational research: Toward a more useful, more influential, and better-funded enterprise. *Educational Researcher, 32*(9), 3-14.

*Callahan, C., & Smith, R. M. (1990). Keller's personalized system of instruction in a junior high gifted program. *Roeper Review, XIII*(1), 39-44.

Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research*. Skokie, IL: Rand McNally.

*Chilton, C. M. (2001). *The math achievement, experiences, and attitudes of gifted and promising math students*. Unpublished doctoral dissertation, Arizona State University.

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. New York: Academic.

Cohen, J. (1977). Statistical power analysis for the behavioral sciences (rev. ed). New York: Academic Press.

Colangelo, N., Assouline, S. G., & Gross, M. U. M. (2004). *A nation deceived*: *How schools hold back America's brightest students, volume II* (pp.13-22). Iowa City, IA: The Connie Belin and Jacqueline N. Blank International Center for Gifted Education and Talent Development.

Colangelo, N., Assouline, S. G., & Lupkowski-Shoplik, A. E. (2004). *Whole-grade acceleration*. In N. Colangelo, S. G. Assouline, & M. U. M. Gross (Eds.), *A nation deceived*: *How schools hold back America's brightest students, volume II* (pp.13-22). Iowa City, IA: The Connie Belin and Jacqueline N. Blank International Center for Gifted Education and Talent Development.

Cook, T., & Campbell, D. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston, MA: Houghton Mifflin.

Cooper, H. (1998). Synthesizing research: A guide for literature reviews (3rd ed.). Beverly Hills, CA: Sage.

Cooper, H. (2007). *Evaluating and interpreting research syntheses in adult learning and literacy*. Harvard Graduate School of Education, National Center for the Study of Adult Learning and Literacy (NCSALL) Occasional Paper. Retrieved from: http://www.ncsall.net/fileadmin/resources/research/op_research_syntheses.pdf

Cooper, H. M., & Hedges, L. V. (Eds.). (1994). *The handbook of research synthesis*. New York: Russell Sage Foundation.

*Cornell, D. G., Callahan, C. M., & Loyd, B. H. (1991). Personality growth of female early college entrants: A controlled, prospective study. *Gifted Child Quarterly, 35,13*5 -143.

Cox, D. R. (1970). *Analysis of binary data*. New York: Chapman & Hall/CRC.

Cumming, G., & Finch, S. (2001). A primer on the understanding, use and calculation of confidence intervals that are based on central and noncentral distributions. *Educational and Psychological Measurement, 61*, 532-575.

Dabrowski, K., & Piechowski, M. M. (1977). *Theory of levels of emotional development*. Oceanside, NY: Dabor.

Daurio, S. P. (1979). Educational enrichment versus acceleration: A review of the literature. In W. C. George, S. J. Cohn, & J. C. Stanley (Eds.), *Educating the gifted: Acceleration and enrichment* (pp. 13–63). Baltimore: Johns Hopkins University Press.

Dooley, D. (2006). *Interpreting Study Conflicts*. University of California, Irvine, Department of Psychology and Social Behavior. Retrieved from: http://www.seweb.uci.edu/users/addooley/StudyConflicts11.18.06.pdf

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

Egger, M., Davey, S. G., & Altman, D. G. (2001). *Systematic reviews in health care: Meta-analysis in context*. London: BMJ Publishing Group.

Eysenck, H. J. (1978). An exercise in mega-silliness. *American Psychologist, 33*, p.517.

Farrington, D. (2003). Methodological quality standards for evaluation research. *The Annals of the American of Political and Social Science, 587*, 49-68.

Fleiss, J. L. (1994). Measures of effect size for categorical data. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 245-260). New York: Russel Sage Foundation.

*Fowler, M. D. (2007). *A program evaluation of achieving a college education plus*. Unpublished doctoral dissertation, Northern Arizona University, AAT 3296083.

*Gatné, F., & Gagnier, N. (2004). The socio-affective and academic impact of early entrance to school. Roeper Review, 26, 128-139.

Glass, G. V. (1976). *Primary, secondary, and meta-analysis of research*. *Educational Researcher, 5*, 3-8.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage

Glasziou, P., Irwig, L., Bain, C., & Goldiz, G. (2001). *Systematic reviews in health care: A practical Guide*. Cambridge: Cambridge University Press.

Gleser, L. J., & Olkin, I. (1994). Stochastically dependent effect sizes. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp.339-355). New York: Sage.

Greenwald, A. G. (1975). Consequence of prejudice against the null hypothesis. *Psychological Bulletin, 82*, 1-20.

Gross, M. U. M. (2003). *Exceptionally gifted children* (2nd ed.) London: Routledge Falmer.

Gross, M. U. M. (2004). In N. Colangelo, S. G. Assouline, & M. U. M. Gross (Eds.), *A nation deceived*: *How schools hold back America's brightest students, volume II* (pp.87-96). Iowa City, IA: The Connie Belin and Jacqueline N. Blank International Center for Gifted Education and Talent Development.

Gross, M. U. M. & Vliet, H. E. van (2005). Radical acceleration and early entry to college: A review of the research. *Gifted Child Quarterly, 49*(2), 154-171.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*, 107-128.

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

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Hedges, L., Shymansky, J., & Woodworth, G. (1989). *A practical guide to modern methods of meta-analysis*. Washington, DC: National Science Teachers Association.

Hedges, L. V., & Waddington, T. (1993). From evidence to knowledge to policy: Research synthesis for policy formation. *Review of Educational Research, 63*(3), 345-352.

Hertberg-Davis, H., & Callahan, C. M. (2007). Advanced placement and international baccalaureate programs. In C. M. Callahan & J. A. Plucker (Eds.), *Critical issues and practices in gifted education: What research says* (pp.31-44). Waco, Texas: Prufrock Press.

*Houston, I. S. (1999). The influences of intentional mentoring on the career aspirations and educational expectations of middle school African American adolescents participating in a college-preparatory program. (Unpublished doctoral dissertation, University of South Carolina).

*Hsu, L. (2003**).** Measuring the effectiveness of summer intensive physics courses for gifted students: A pilot study and agenda for research**.** *Gifted Child Quarterly, 47*, 212-218.

Hunter, E. J., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.

Hunter, J. H., Schmidt, F. L., & Jackson, G. B. (1982). *Meta-analysis: Accumulating research findings across studies*. Beverly Hills, CA: Sage.

*Ingersoll, K. S. (1992). *Social adjustment of female early college entrants*. Unpublished doctoral dissertation, University of Virginia, AAT 9237472.

*Janos, P. M. (1987). A fifty-year follow-up of Terman's youngest college students and IQ matched agemates. *Gifted Child Quarterly, 31*(2), 55-58.

*Janos, P. M., & Robinson, N. M. (1985). The performance of students in a program of radical acceleration at the university level. *Gifted Child Quarterly, 29*(4), 175-179.

*Janos, P. M., Robinson, N. M., & Lunneborg, C. E. (1989). Markedly early entrance to college: A multi-year comparative study of academic performance and psychological adjustment. *Journal of Higher Education, 60*, 496-518.

*Jin, S.-U., & Moon, S. M. (2006). A study of well-being and school satisfaction among academically talented students attending a science high school in Korea. *Gifted Child Quarterly, 2006, 50,* 169 - 184.

Kalaian, H. A., & Raudenbush, S. W. (1996). A multivariate mixed linear model for meta-analysis. *Psychological Methods, 1*, 227-235.

Kavale, K. A. (1984). Potential advantages of the mate-analysis techniques for research in special education. *The Journal of Special Education, 18*, 61-72.

Kavale, K. A., & Glass, G. V. (1984). Meta-analysis and policy decisions in special education. In B. K. Keogh (Ed.), *Advances in special education* (Vol. IV. Pp. 195-247). Greenwich, CT: JAI.

Kelley, P. J. (2007). A meta-analysis of teacher education effects in center-based early childhood education classrooms. Unpublished doctoral dissertation, Columbia University, New York.

Kent, S. D. (1992). The effects of acceleration on the social and emotional development of gifted elementary students: A meta-analysis. *Dissertation Abstracts International, 54*(2-A). U.S.: University Microfilms International.

Kulik, J. A. (1991). *Ability grouping*. Research-based decision-making series. Storrs, CT: National Research Center on the Gifted and Talented, University of Connecticut.

Kulik, J. A. (1993). An analysis of the research on ability grouping. *NRC/GT Newsletter*, *Spring*, 8–9.

Kulik, J. A. (2004). Meta-analytic studies of acceleration. In N. Colangelo, S. G. Assouline, & M. U. M. Gross (Eds.), *A nation deceived*: *How schools hold back America's brightest students, volume II* (pp.13-22). Iowa City, IA: The Connie Belin and Jacqueline N. Blank International Center for Gifted Education and Talent Development.

Kulik, J. A., & Kulik, C. C. (1984). Effects of accelerated instruction on students. *Review of Educational Research, 54*(3), 409-425.

Kulik, J. A., & Kulik, C.-L. C. (1989). Meta-analysis in educational research [monograph]. *International Journal of Educational Research, 13*, 221-340.

Kulik, J. A., & Kulik, C.-L. C. (1991). Ability group and gifted students. In N. Colangelo & G. Davis (Eds.), *Handbook of gifted education* (pp. 178-196). Boston, MA: Allyn & Bacon.

Kulik, J. A., & Kulik, C. C. (1992). Meta-analytic findings on grouping programs. *Gifted Child Quarterly, 36*, 73-77.

Lambert, R. (1995). The robustness of the standard errors of summarized, corrected validity coefficients to non-independence and non-normality of primary data. Unpublished doctoral dissertation, Georgia State University, Atlanta, GA.

Lauer, J., & Asher, J. W. (1988). *Composition research: Empirical designs*. New York: Oxford University Press.

Leandro, G. (2005). *Meta-analysis in medical research*. Blackwell Publishing, BMJ Books.

Lee, S.-Y., & Olszewski-Kubilius, P. (2005). Investigation of high school credit and placement for summer coursework taken outside of local schools. *Gifted Child Quarterly, 49*(1), 37-50.

Lee, S.-Y., & Olszewski-Kubilius, P. (2006). Comparisons between talent search students qualifying via scores on standardized tests and via parent nomination. *Roeper Review*, *28*(3), 157-166.

Light, R. J., & Pillemer, D. B. (1984). *Summing up: The science of reviewing research*. Cambridge: Harvard University Press.

Light, R. J., & Smith, P. V. (1971). Accumulating evidence: Procedures for resolving contradictions among different studies. *Harvard Educational Review, 41*, 429-471.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage Publications.

Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B., & d□Apollonia, S. (1996). Within-class grouping: A meta-analysis. *Review of Educational Research, 66*(4), 423-458.

Lubinski, D. (2004). In N. Colangelo, S. G. Assouline, & M. U. M. Gross (Eds.), *A nation deceived*: *How schools hold back America's brightest students, volume II* (pp.23-37). Iowa City, IA: The Connie Belin and Jacqueline N. Blank International Center for Gifted Education and Talent Development.

Lubinski, D., Benbow, C. P., Shea, D. L., Eftekhari-Sanjani, H., & Halvorson, M. B. J. (2001). Men and women at promise for scientific excellence: Similarity not dissimilarity. *Psychological Science, 12*, 309-317.

Lubinski, D., Webb, R. M., Morelock, M. J., & Benbow, C. P. (2001). Top 1 in 10,000: A-10 year follow-up of the profoundly gifted. *Journal of Applied Psychology, 86*, 718-729.

*Lupkowski, A. E., Whitmore, M., & Ramsay, A. (1992). The impact of early entrance to college on self-esteem: A preliminary study. *Gifted Child Quarterly. Special Issue: Challenging the Gifted: Grouping and Acceleration, 36*(2), 87-90.

*Ma, X. (2002). Early acceleration of mathematics students and its effect on growth in self-esteem: A longitudinal study. *International Review of Education, 48*(6), 443-468.

*Ma, X. (2005). A Longitudinal Assessment of Early Acceleration of Students in Mathematics on Growth in Mathematics Achievement. *Developmental Review, 25* (1), 104-131.

McCoach, D. B., Kehle, T. J., Bray, M. A., & Siegle, D. (2001). Best practices in the identification of gifted students with learning disabilities. *Psychology in the Schools, 38*(5), 403-411.

*Mills, C. J., Ablard, K. E., & Gustin, W. C. (1994). Academically talented students' achievement in a flexibly paced mathematics program. *Journal for Research in Mathematics Education, 25*, 495-511.

*Moon, T. R., & Callahan, C. M. (2001). Curricular modifications, family outreach, and a mentoring program: Impacts on achievement and gifted identification in high-risk primary students. *Journal for the Education of the Gifted, 24*(4), 305-321.

Moon, S. M., & Reis, S. M. Acceleration and twice-exceptional students (2004). In N. Colangelo, S. G. Assouline, & M. U. M. Gross (Eds.), *A nation deceived*: *How schools hold back America's brightest students, volume II* (pp.109-119). Iowa City, IA: The Connie Belin and Jacqueline N. Blank International Center for Gifted Education and Talent Development.

National Association for Gifted Children (November, 1992). Position Paper on Acceleration. Washington, DC.

National Research Council (1999). *Improving student learning: A strategic plan for education research and its utilization*. Committee on Feasibility Study for a Strategic Education Research Program. Washington DC: National Academy Press.

National Research Council (2002a). *Scientific Research in Education*. Committee on Scientific Principles for Education Research, Richard J. Shavelson and Lisa Towne. Washington DC: National Academies Press.

National Research Council (2002b). Advancing *Scientific Research in Education*. Committee on Research in Education, Lisa Towne, Lauress L. Wise, and Tina M. Winters. Washington DC: National Academies Press.

Neihart, M. (2007). The socioaffective impact of acceleration and ability grouping: Recommendations for best practice. *Gifted Child Quarterly, 51*, 330-341.

Olenchak, F. R. (1994).Talent development: Accommodating the social and emotional needs of secondary gifted/learning disabled students. *Journal of Secondary Gifted Education, 5*, 40-52.

Olenchak, F. R. (1995). Effects enrichment on gifted/learning disabled students. *Journal for the Education of the Gifted, 18*, 385-399.

Olenchak, F. R., & Reis, S. M. (2001). Gifted students with learning disabilities. In M. Neihart & S. M. Reis & N. M. Robinson & S. M. Moon (Eds.), *The social and emotional development of gifted children* (pp.267-289). Waco, TX: Prufrock Press.

Olszewski-Kubilius, P. (1995). A summary of research regarding early entry to college. *Roeper Review, 18*, 121-125.

Olszewski-Kubilius, P. (2004). In N. Colangelo, S. G. Assouline, & M. U. M. Gross (Eds.), *A nation deceived*: *How schools hold back America's brightest students, volume II* (pp.69-76). Iowa City, IA: The Connie Belin and Jacqueline N. Blank International Center for Gifted Education and Talent Development.

Pierce, C. A. (2007). Software Review: Review of Borenstein et al.'s (2006) *Comprehensive Meta-Analysis* (v. 2.2.027) 10.1177/1094428106296641. *Organizational Research Methods OnlineFirst*, published on July 27, 2007 as doi:10.1177/1094428106296641, from http://orm.sagepub.com/cgi/rapidpdf/1094428106296641v1.pdf

*Plucker, J. A., & Taylor, J. W., V. (1998). Too much too soon? Non-radical advanced grade placement and the self-concept of gifted students. *Gifted Education International, 13*(2), 121-135.

Pollins, L. D. (1983). The effects of acceleration on the social and emotional development of gifted students. In C. P. Benbow & J. C. Stanley (Eds.), *Academic precocity: Aspects of its development* (pp. 160-178). Baltimore, MD: Johns Hopkins University Press.

Pressey, S. L. (1949). *Educational acceleration: Appraisals and basic problems*. Columbus, OH: The Ohio State University.

Reis, S. M., Burns, D. E., & Renzulli, J. S. (1992). *Curriculum compacting: The complete guide to modifying the regular curriculum for high ability students*. Mansfield Center, CT: Creative Learning Press.

*Reis, S. M., Eckert, R. D., McCoach, B., Jacobs, J. K., & Coyne, M. (2008). Using enrichment reading practices to increase reading fluency, comprehension, and attitudes. Journal of Educational Research.

Reis, S. M., Neu, T. W., & McGuire, J. M. (1995). *Talents in two places: Case studies of high ability students with learning disabilities who have achieved*. Storrs, CT: National Research Center on the Gifted and Talented.

*Reis, S. M., Westberg, K. L., Kulikowich, J. M., & Purcell, J. H. (1998). Curriculum compacting and achievement test scores: What does the research say? *Gifted Child Quarterly, 42*, 123-129.

Renzulli, J. S., Smith, L. H., & Reis, S. M. (1982). Curriculum compacting: A essential strategy for working with gifted students. *The Elementary School Journal, 82*(3), 185-194.

*Reynolds, R. A. (1993). *The relationship between early school entrance and eighth-grade school success*. Unpublished doctoral dissertation, University of South Florida, AAT 9323696.

Rich, R. (1983). Knowledge synthesis and problem solving. In S. A. Ward & L. J. Reed (Eds.), *Knowledge structure and use: Implications for synthesis and interpretation* (pp. 285-312). Philadelphia: Temple University Press.

*Richards, J. H. (2006). *College-level learning in high school: Academic achievement comparisons between CLLHS-accelerated and non-accelerated peers in college classes*. Unpublished doctoral dissertation, The University of Kansas, AAT 3206127.

Robinson, N. M. (2004). Effects of academic acceleration on the social-emotional status of gifted students. In N. Colangelo, S. G. Assouline, & M. U. M. Gross (Eds.), *A nation deceived*: *How schools hold back America's brightest students, volume II* (pp.59-67). Iowa City, IA: The Connie Belin and Jacqueline N. Blank International Center for Gifted Education and Talent Development.

*Robinson, N. M., & Janos, P. M. (1986). Psychological adjustment in a college-level program of marked academic acceleration. *Journal of Youth and Adolescence, 15*(1), 51-60.

Roger, K. B. (1991). A best-evidence synthesis of research on types of accelerative programs for gifted students. *Dissertation Abstracts International, 52*, 796-A. (University Microfilms No. AAT91-22206).

Rogers, K. B. (1992). A best-evidence synthesis of the research on acceleration options for gifted learners. In N. Colangelo, S.G. Assouline, & D. L. Ambroson (Eds.), *Talent development: Proceedings from the 1991 Henry B. and Jocelyn Wallace national symposium on talent development* (pp. 406-409). Unionville, NY: Trillium Press.

Rogers, K. B. (2002). *Re-forming gifted education: How parents and teachers can match the program to the child*. Scottsdale, AZ: Great Potential Press.

Rogers, K. B. (2004). The academic effects of acceleration. In N. Colangelo, S. G. Assouline, & M. U. M. Gross (Eds.), *A nation deceived*: *How schools hold back America's brightest students, volume II* (pp.47-57). Iowa City, IA: The Connie Belin and Jacqueline N. Blank International Center for Gifted Education and Talent Development.

Rogers, K. B. (2007). Lessons learned about educating the gifted and talented: A synthesis of the research on educational practices. *Gifted Child Quarterly, 51*, 382-396.

Rogers, K. B., Young, M., &Lonergan, R. (2008). A best-evidence synthesis of the research on academic acceleration 1990-present. Paper presented at The Ninth Biennial Henry B. & Jocelyn Wallace National Research Symposium on Talent Development, Iowa City, IA.

Rosenberg, M. S. (2005). The file-drawer problem revisited: A general weighted method for calculating fail-safe numbers in meta-analysis. *Evolution, 59*, 464-468.

Rosenberg, M. S., Adams, D. C., Gurevitch, J. (2000). *MetaWin: Statistical software for meta-analysis version 2*. Sunderland, Massachusetts: Sinauer Associates.

Rosenthal, R. (1979). The file-drawer problem and tolerance for null results. *Psychological Bulletin, 86*, 638-641.

Rosenthal, R. (1991). *Meta-analytic procedures for social research* (Rev.ed.). Newbury Park, CA: Sage.

Rosenthal, R., & Rosnow, R. L. (1984). *Essentials of behavioral research: Methods and data analysis.* New York: McGraw-Hill.

Sanchez-Meca, J., Marin-Martinez, F., & Chacon-Moscoso, S. (2003). Effect-size indices for dichotomous outcomes in meta-analysis. *Psychological Methods, 8*(4), 448-467.

*Sayler, M. (2008).The effects of acceleration on personal thriving: Psychological well being and academic and personal successes of gifted early college entrants.

*Sayler, M. F., & Brookshire, W. K. (1993). Social, emotional, and behavioral adjustment of accelerated students, students in gifted classes, and regular students in eighth grade. *Gifted Child Quarterly, 37*, 150-154.

Shadish, W. R. Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

Sipe, T. A., & Curlette, W. L. (1997). A meta-synthesis of factors related to educational achievement: A methodological approach to summarizing and synthesizing meta-analyses. *International Journal of Educational Research, 25*, 583-698.

Slavin, R. E. (1986). Best evidence synthesis: An alternative to meta-analytic and traditional reviews. *Educational Researcher, 15*, 5-11.

Slavin, R. E. (1987). Achievement effects of ability grouping in elementary schools: A best-evidence synthesis. *Review of Educational Research, 60*, 293-336.

Slavin, R. E. (1990). Achievement effects of ability grouping in secondary schools: A best-evidence synthesis. *Review of Educational Research, 60*, 471-499.

Slavin, R. E. (1993). Ability grouping in the middle grades: Achievement effects and alternatives. *The Elementary School Journal, 93*(5), Special Issue: Middle Grades Research and Reform, 535-552.

Southern, W. T., & Jones, E. D. (Eds.) (1991). *The academic acceleration of gifted children*. New York: Teachers College Press.

Southern, W. T., & Jones, E. D. (2004). In N. Colangelo, S. G. Assouline, & M. U. M. Gross (Eds.), *A nation deceived*: *How schools hold back America's brightest students, volume II* (pp.5-12). Iowa City, IA: The Connie Belin and Jacqueline N. Blank International Center for Gifted Education and Talent Development.

Southern, W. T., Jones, E. D., & Stanley, J. C. (1993). Acceleration and enrichment: The context and development of program options. In K. A. Heller, F. J. Mönks, & A. H. Passow (Eds.), *International handbook of research and development of giftedness and talent* (pp. 387–405). New York: Pergamon.

*Stamps, L. (2004). The effectiveness of curriculum compacting in first grade classrooms. *Roeper Review, 27*, 31-41.

Stanley, J. C. (1978). Radical acceleration: Recent educational innovation at JHU. *Gifted Child Quarterly, 22*, 62-67.

Stanley, J. C., Keating, D. P., & Fox, L. H. (Eds.). (1974). *Mathematical talent: Discovery, description, and development*. Baltimore: Johns Hopkins University Press.

Sterling, T. D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance-or vice versa. *Journal of the American Statistical Association, 54*, 30-34.

Strube, M. J. (1991). Multiple determinants and effect size: A more general method of discourse. *Journal of Personality and Social Psychology, 61*, 1024-1027.

Sutton, A. J., Abrams, K. R., Jones, D. R., Sheldon, T. A., & Song, F. (2000). *Methods for meta-analysis in medical research*. Chichester: Wiley.

Swiatek, M. A. (2007). The talent search model: Past, present, and future. *Gifted Child Quarterly, 51*, 320-329.

*Swiatek, M. A., & Benbow, C. P. (1991a). Ten-year longitudinal follow-up of ability-matched accelerated and unaccelerated gifted students. *Journal of Educational Psychology, 83*, 528-538.

*Swiatek, M.A., & Benbow, C. P. (1991b). A ten-year longitudinal follow-up of participants in a fast-paced Mathematics course. *Journal for Research Mathematics Education, 22*, 138-150.

Tannenbaum, A. J. (1958). History of interest in the gifted. In N. Henry (Ed.), *Education for the gifted*. 57th Yearbook of the National Society for the Study of Education (Part II, pp.21-38). Chicago: University of Chicago Press.

Tellegen, A., Ben-Porath, Y.S., McNulty, J.L., Arbisi, P.A., Graham, J.R., & Kaemmer, B. (2003). *The MMPI-2 Restructured Clinical Scales: Development, validation, and interpretation*. Minneapolis, MN: University of Minnesota Press.

Terman, L. M., & Oden, M. H. (1959). *Genetic studies of genius. IV. The gifted group at midlife: Thirty-five years 'follow-up of the superior child.* Stanford, CA: Stanford University Press.

Thompson, B. (1999). Improving research clarity and usefulness with effect size indices as supplements to statistical significance tests. *Exceptional Children, 65*, 329-338.

Thompson, B. (2002). What future quantitative social science research could look like: Confidence intervals for effect sizes. *Educational Researcher, 31*(3), 24-31.

Thompson, B. (2006). Research synthesis: Effect sizes. In J. Green, G. Camilli, & P. B. Elmore (Eds.), *Complementary methods for research in education*. Washington, DC: American Educational Research Association.

United States Department of Education (2002). *Education Sciences Reform Act of 2002*. Washington, DC: The Senate and House of Representatives of the United States of America in Congress Assembled.

United States Department of Education (2008). *Jacob K. Javits Gifted and Talented Student Education*. Washington, DC: Office of Elementary and Secondary Education (OESE)

United States Department of Education (1993). *National excellence: A case for developing America's talent*. Washington, DC: United States Government Printing Office

Vacha-Haase, T. & Thompson, B. (2004). How to estimate and interpret various effect sizes. *Journal of Counseling Psychology, 51*, 473-481.

VanTassel-Baska, J. (1992). Educational decision making on acceleration and grouping. *Gifted Child Quarterly, 36*, 68-72.

Vaughn, V. L., Feldhusen, J., & Asher, J. W. (1991). Meta-analyses and review of research on pull-out programs in gifted education. *Gifted Child Quarterly, 35*(2), 92-98.

Vockell, E. L., & Asher, J. W. (1995). *Educational research* (2nd ed.). Englewood Cliffs, NJ: Macmillan.

Walberg, H. J., & Haertel, E. H. (1980). Research integration: An introduction and overview. *Evaluation in Education, 4*, 5-10.

Wang, M. C., & Bushman, B. J. (1999). *Integrating results through meta-analytic review using SAS software*. Cary, NC: SAS Institute Inc.

*Washington, J. (1999). *Achievement and social-emotional outcomes of a multigraded, accelerated mathematics program*. Unpublished doctoral dissertation, The Johns Hopkins University, AAT 9964225.

*Weiner, D. A. (1985). The role of mentors in the growth and development of gifted children and adolescents. *Dissertation Abstracts International, 4446* (9-A), 2639.

*Wells, R., Lohman, D., &Marron, M. (2008). What factors are associated with grade acceleration? An analysis and comparison of two U. S. databases.

What Works Clearinghouse (2007, April 16). Technical details of WWC-conducted computations. Retrieved April 1, 2008, from http://ies.ed.gov/ncee/wwc/pdf/conducted_computations.pdf

White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin, 91*, 461-481.

Wollf, F. M. (1986). *Meta-analysis: Quantitative methods for research synthesis*. Beverly Hills, CA: Sage Publications.

Wortman, P.M., (1994). Judging research quality. In Cooper, H., & Hedges, L.V. (Eds.) 1994. *The handbook of research synthesis* (pp. 97-110). New York: Russell Sage Foundation.

*Ysseldyke, J., Tardrew, S., Betts, J., Thill, T., & Hannigan, E. (2004). Use of an instructional management system to enhance math instruction of gifted and talented students. *Journal for the Education of the Gifted, 27*(4), 293-319.

*Studies marked with an asterisk in the reference list were included in the meta-analysis.

APPENDICES

Appendix A. Study Variables Grouped by Social Development and Emotional Development

Social Development Factors

Indicators of Participation

Club Membership	Club Officer	Participation in Groups
Work Spirit	Group Spirit	Out-of-School Activity
Group Conformity	Leadership	Extra-Curricular
Capacity for Status	Leisure Activity	Social-Recreational

Indicators of Relationship

Family Relations	Communality	Friendship Choices
Good Impression	Communication	Community Relations
Places Liked Best	Social	Teacher Rating
Social Presence	Social Skills	Verbal Interaction
School Relations	Social Standing	Personal/Psychological
Sociability	Socialization	Social/Psychological
Social Handicap*	Self-Centered*	Anti-Social Tendencies*

*high scores on these variables were considered negative effects

Emotional Development Factors

Indicators of Academic Inclination

Academic Orientation	Intellectual Efficiency
Achievement via Independence	Learning
Achievement via Conformity	Poor Work Habits*
Lack of Intellectual Independence*	

Indicators of Behavioral

Behavior Rating	School-Related Behavior
Anxious Producer*	Nervous Symptoms*
Verbal Negativism*	Classroom Disturbance*
Disturbance-Restlessness*	

Indicators of Attitude

Attitude Toward Math	General School Interest
Attitude Toward School	Affect of Acceleration
Attitude	General Affect

Indicators of Autonomy

Responsibility	Self Confidence/Concept
Self-Reliance	Socieometric Choices
Personal Worth	Emotional/Social Maturity
Personal Freedom	Emotional Stability
Self-Acceptance	Locus of Control
Self-Esteem	Self-Adjustment
Well Being	Rational Conscientiousness
Self-Control	Irrational Conscientiousness*
External Blame*	Expressed Inability*

Indicators of Analytical

Planning	Originality	Creativity
Investigative	Reasoning	Risk-Taking
Flexibility		

Indicators of Personality

Psychological Mindedness	Belonging	Motivation
Introvert/Extrovert	Initiative	Femininity
Conventional	Tolerance	Economic
Theoretical	Realistic	Enterprising
Artistic	Political	Religious

Quiet/Withdrawn*	Withdrawn*	Dominance*
General Anxiety*	Personal Adjustment	
Dogmatic/Inflexible*	Aesthetic	

*high scores on these variables were considered negative effects

Adopted from Kent (1992, pp. 77-79)

Appendix B. Study Feature Coding Descriptors

Item	Item Name	Descriptions
		General Features
1	Study ID	Identification number of the study
2	Study Name	e.g., Swiatek & Benbow, 1991
3	Study Focus	1.Academic achievement 2. Social-emotional development 3. Both
4	Identification	Identification of high-ability learners 1. Achievement scores 2. IQ 3. Age 4. Teacher/Parent identification 5. Self-reported 6. Multiple ways 7. Enrolled in gifted/enrichment program/class 8. Classified as G/T in regional, such as states, district, schools, etc 9. Qualified for certain class, school, college, or grade level
5	Acceleration categories I	1. Grade-based acceleration 2. Subject-based acceleration 3. Both 1 and 2 4. Non-specific

Item	Item Name	Descriptions
6	Acceleration categories II	1. Radical acceleration 2. Non-radical acceleration 3. Both 1 and 2 4. Non-specific
7	Acceleration forms	The most common acceleration forms: 1. Early kindergarten, school or college entrance 2. Grade skipping 3. Advanced placement/Dual Credits/ International Baccalaureate 4. Subject-matter acceleration (e.g., math and reading) 5. Curriculum compacting 6. Multiple forms 7. Early graduation 8. Mentoring
8	Acceleration Duration	1.Under 14 weeks 2. 14 to 28 weeks 3. Over 28 weeks 4. Non-specific
9	Treatment and comparison groups	The description of treatment and comparison groups, such as accelerated gifted students are in treatment groups, while non-accelerated gifted students construct comparison groups. Or, more accelerated gifted students are categorized as treatment group, and less accelerated gifted students are used as comparison groups.
10	Treatment group categories	1. Accelerated gifted students 2. Comparatively more accelerated gifted students 3. Both 1 and 2

Item	Item Name	Descriptions
11	Comparison group categories	1. Non accelerated gifted students 2. Comparatively less accelerated gifted students 3. Both 1 and 2
12	Comparison Perspectives	1. Accelerants VS. non-accelerants of equivalent aptitudes and abilities, with same-age peers 2. Accelerants VS. non-accelerants of equivalent aptitudes and abilities, with older-age peers 3. More accelerated students VS. less accelerated students 4. Both 1 and 2 5. Both 1 and 3 6. Both 2 and 3 7. Accelerants vs. Non-accelerants mixed-age peers
13	Grade Level at Acceleration	Student grade level when experienced acceleration: 1. Elementary 2. Middle school 3. High school 4. Post secondary 5. Both 1 and 2 6. Both 2 and 3 7. Both 3 and 4 8. Kindergarten 9. varying (include multiple levels)

Item	Item Name	Descriptions
14	Mean age at Acceleration	The approximate or exact mean age of treatment group at the beginning of acceleration 1. ≤ 7 years old 2. Between 8 to 11 years old 3. Between 12 to 17 years old 4. between 18 to 35 years old 5. Both 1 and 2 6. Both 2 and 3 7. Both 3 and 4 8. Varying
15	Mean age of treatment group at data collection	1. ≤ 7 years old 2. Between 8 to 11 years old 3. Between 12 to 17 years old 4. between 18 to 35 years old 5. Varying (for example, subjects in longitudinal studies)
16	Mean age of comparison group at data collection	1. ≤ 7 years old 2. Between 8 to 11 years old 3. Between 12 to 17 years old 4. between 18 to 35 years old 5. Varying (for example, subjects in longitudinal studies)
17	Study Time	1. Study conducted within 1 year after acceleration 2. Longitudinal/retrospective studies 3. Both

Item	Item Name	Descriptions
18	School type	1. Rural 2. Suburban 3. Urban 4. Not applicable 5. National
19	SES	Average socioeconomic status of students in treatment and comparison groups. 1. Low 2. Medium 3. High 4. Non-specific 5. Low to medium 6. Medium to high
20	Ethnicity	The Partitioning of the sample by Ethnicity 1. Non-minority students dominant (over 60%) 2. Minority students dominant (over 60%) 3. Approximately equivalent 4. Non-specific
21	Gender	Sex ratio of treatment groups: 1. Male predominant (over 60%) 2. Female predominant (over 60%) 3. Approximately equivalent 4. Non-specific

Item	Item Name	Descriptions
		Methodological Features
22	Study retrieval source	1.Journal article 2.Doctoral dissertation 3.Unpublished study obtained from author
23	Sample Assignment	Method of sample assignment to treatments: 1. Randomized 2. Non-randomized 3. Mixed
24	Comparison group construction	The approaches of constructing comparison groups: 1. Matching IQ, achievement scores, age, or grade-level. 2. Ex post facto design (relying on archival data to construct comparison groups) 3. Randomization from initial pool of eligible subjects 4. Randomization from clustered sample groups 5. Self-reported information 6. The rest of eligible subjects after treated groups have been assigned.
25	Subject group number	Total subject group numbers (include both treatment and comparison groups) 1. Two groups 2. More than two groups
26	Test design	Type of test designs for measuring the study outcomes: 1.Posttest only 2. Pretest-posttest 3. Multiple times tests
27	Research design	Type of research designs: 1. Experimental 2. Quasi-experimental 3. Causal-comparative research

Item	Item Name	Descriptions
28	Sample Size	Total sample size categories(include both treatment and comparison groups) 1. ≤ 100 2. 101-500 3. 501-1000 4. >1000 5. Not applicable (e.g., didn't report the sample size for both treatment and comparison group, or results came from not appropriate unit of analysis)
29	Statistical Analysis	Study outcomes from statistical analysis procedures [a]: 1. Descriptive statistics (including dichotomous outcomes) 2. Student-level T-test, ANOVA, or MANOVA 3. Student-level ANCOVA, MANCOVA, Logistic Regression Analysis 4. Cluster-level analysis 5. Hierarchical Linear Model (HLM) analysis with cluster-level assignment
30	Statistical power	Statistical power calculable from the study: 1. Yes 2. No
31	Data source	1. Standardized tests 2. Non-standardized tests 3. Surveys/questionnaires 4. Combing 1, or 2, or 3, or 5. 5. Accumulative records or archival data 6. National databases

Item	Item Name	Descriptions
32	Test type	1. Standardized tests 2. Ad hoc written testing(test performed without planning and documentation) 3. Criterion-referenced test 4. Researcher/teacher developed scales/tests 5. Accumulative records 6. Other (such as one-time survey) 7. Not-applicable 8. Combing 1, 2, or 3
33	Test subject	1. Science 2. Technology 3. Engineering 4. Mathematics 5. Humanities 6. Other (student attitudes/beliefs) 7. Multiple (including reading, math, science, social science, attitudes, etc.) 8. Not-applicable
34	Reliability information	Reliability information provided: 1. Yes 2. No
35	Measurement reliability Information	Measurement reliability information provided 1. Yes 2. No

Item	Item Name	Descriptions
36	Measurement reliability status	1. High reliability (e.g. standardized tests, national databases) 2. Moderate reliability(e.g., researcher developed instruments, scales with moderate measurement reliability) 3. Low reliability(e.g., researcher/teacher developed tests or measurements without solid measurement information) 4. Not-available
37	Study validity status [b]	1. Low validity (average study validity is between 0 and 2.0) 2. Moderate validity (average study validity is between 2.1 and 3.0) 3. High validity (average study validity is between 3.1 and 4.0)
38	ES reporting	Effect size reported in the study by primary study author(s) 1. Yes 2. No
39	Effect Size Extraction	1. Calculated from descriptive statistics 2. Calculated from inferential statistics 3. Estimated from probabilities 4. Computed effect sizes in original study
		Other
40	Page number	Page number where the data for effect size computation is retrieved.
41	Coding time	The coding time to finish each study

[a]This method of category is adopted from the WWC (2007).
[b]Study validity in this meta-analysis denotes five kinds of validity, as suggested by Kelley (2007). They are: internal validity, construct validity, statistical validity, external validity, and descriptive validity. This meta-analysis uses a working sheet, adopted from Kelley (2007) to measure the study validity (see Appendix B Study Validity Measuring Sheeting).

Appendix C. Study Validity Measuring Sheet

Study ID#__________

1. Internal Validity

A. Design
   i. Randomized, controlled, double-blind trials
   ii. Quasi-experimental studies (experiments without randomization)
   iii. Controlled observational studies (comparison of outcomes between participants who have received an intervention and those who have not)
   iv. Observational studies without a control group
   v. Expert opinion

B. Selection: occurs when preexisting difference between experimental and control conditions influence effects.

C. History: the effect is caused by some external event occurring at the same time as the intervention.

D. Maturation: A preexisting trend (e.g., normal human development) influences the effect.

E. Instrumentation: the observed effect is caused by a change in the method used to measure the outcome.

F. Testing: the pretest measurement causes a change in the posttest measure.

G. Regression to the mean: when an intervention is implemented on units with unusually high scores, natural fluctuation will cause a decrease in these scores on the posttest, which may be mistakenly interpreted as an effect of the intervention. This effect can occur in the opposite direction, resulting in an increase in scores on posttest when interventions are applied to low-scoring groups.

H. Differential attrition: effects are influenced by differential loss of subjects from the experimental compared to control conditions.
I. Causal order: uncertainty with regard to whether the intervention preceded the outcome.
J. Mediating or moderating effects

Final Score: 0 1 2 3 4

2. Construct Validity
   i. Extent to which the intervention succeeded in changing what it was intended to change (e.g., treatment fidelity or implementation failure).
   ii. Validity and reliability of outcome measures.
   iii. Participant's knowledge of the intervention (e.g., the "Hawthorn Effect").
   iv. Contamination of treatment (e.g., where the control group receives elements of the intervention).
   v. Potential interaction effects between different interventions or different ingredients of an intervention.

Final Score: 0 1 2 3 4

3. Statistical Conclusion Validity
   i. Statistical significance calculated?
   ii. Effect size calculated?
   iii. Insufficient statistical power to detect the effect (e.g., due to small sample size).
   iv. Use of inappropriate statistical techniques (e.g., where data violate the underlying assumptions of a statistical test).

v. Use of many statistical tests ("fishing" for significant results).

vi. Random heterogeneity across respondents, setting, or persons implementing the treatment.

Final Score: 0 1 2 3 4

4. External Validity:

   i. Sample generalizability

Final Score: 0 1 2 3 4

5. Descriptive Validity

   A. Study Design: how were experimental units allocated to experimental or control conditions.
   B. Demographic characteristics of experimental units and settings: for example, age and gender of participants, sociodemographic features of sites or areas.
   C. Sample size and attrition rates.
   D. Description of causal hypotheses to be tested and theories from which they are derived.
   E. Operational definition and detailed description of the intervention, including duration and intensity.
   F. Implementation details and program delivery personnel.
   G. Description of treatment received by control group.
   H. Operational definition and measurement of the outcome before and after the intervention.

I. Reliability and validity of outcome measures
J. Details of follow-up period after the intervention.
K. Effect size, confidence intervals, statistical significance, and statistical methods used.
L. How were independent and extraneous variables controlled and how were other threats to internal validity ruled out?
M. Conflict of interest: who funded the intervention, how independent were the researchers?

Final Score: 0 1 2 3 4

Note: This coding sheet was adopted from Kelley (2007) with minor revision.

Appendix D. Academic Achievement Study Outcomes and Effect Sizes Extracted

No.	**Study**	**Subgroup within study**	**Comparison group**	**Outcome variable**	**Hedges's *g***	**Variance**
1	Janos & Robinson, 1985	Post-secondary Achievement	older age peers	1. University credits earned	0.014	0.081
2	Janos & Robinson, 1985	Post-secondary Achievement	older age peers	2. College GPA	0.536	0.085
3	Janos & Robinson, 1985	Post-secondary Achievement	older age peers	3. The Concept Mastery Test	0.310	0.082
4	Brody & Benbow, 1987	P-12 Achievement	mixed-age peers	1. SAT-M	0.620	0.020
5	Brody & Benbow, 1987	P-12 Achievement	mixed-age peers	2. SAT-V	0.307	0.031
6	Brody & Benbow, 1987	P-12 Achievement	mixed-age peers	3. Talent search SAT-M	0.555	0.018
7	Brody & Benbow, 1987	P-12 Achievement	mixed-age peers	4. Talent search SAT-V	0.242	0.018
8	Brody & Benbow, 1987	P-12 Achievement	mixed-age peers	5. College Board Achievement Tests	0.175	0.030
9	Brody & Benbow, 1987	P-12 Achievement	mixed-age peers	6. GPA	0.199	0.018
10	Brody & Benbow, 1987	P-12 Achievement	mixed-age peers	7.Total awards	0.170	0.017

No.	Study	Subgroup within study	Comparison group	Outcome variable	Hedges's *g*	Variance
11	Brody & Benbow, 1987	P-12 Achievement	mixed-age peers	8. National and State Awards	0.251	0.017
12	Brody, Assouline, & Stanley, 1990	P-12 Achievement	older age peers	1. SAT-M	0.589	0.016
13	Brody, Assouline, & Stanley, 1990	P-12 Achievement	older age peers	2. SAT-V	0.613	0.016
14	Brody, Assouline, & Stanley, 1990	P-12 Achievement	older age peers	3. Attended public high school	0.242	0.025
15	Brody, Assouline, & Stanley, 1990	P-12 Achievement	older age peers	4. Earning AP credit	0.299	0.021
16	Brody, Assouline, & Stanley, 1990	Post-secondary Achievement	older age peers	1. Graduated from college 4 years or less	2.237	0.040
17	Brody, Assouline, & Stanley, 1990	Post-secondary Achievement	older age peers	2. graduated from college 3.5 years or less	0.682	0.026
18	Brody, Assouline, & Stanley, 1990	Post-secondary Achievement	older age peers	3. Earning concurrent BA/MA	1.383	0.063
19	Brody, Assouline, & Stanley, 1990	Post-secondary Achievement	older age peers	4. Graduated with general honors	0.426	0.022
20	Brody, Assouline, & Stanley, 1990	Post-secondary Achievement	older age peers	5. Graduated with departmental honors	0.575	0.024

No.	Study	Subgroup within study	Comparison group	Outcome variable	Hedges's *g*	Variance
21	Brody, Assouline, & Stanley, 1990	Post-secondary Achievement	older age peers	6. Graduated with Phi Beta Kappa or Tau Beta Pi	0.521	0.029
22	Sayley & Brookshire, 1993	P-12 Achievement	older age peers	Composite score of achievement test	0.081	0.006
23	Hsu, 2003	P-12 Achievement	mixed-age peers	The effectiveness of summer intensive physics course	-2.493	0.013
24	Callahan & Smith, 1990	P-12 Achievement	same age peers	the effects of self-paced psychology class	1.671	0.261
25	Noble & Robinson, 1993	Post-secondary Achievement	older age peers	1. Undergraduate Grade Point Average	0.054	0.047
26	Noble & Robinson, 1993	Post-secondary Achievement	older age peers	2. Total BA earned	0.103	0.060
27	Noble & Robinson, 1993	Post-secondary Achievement	older age peers	3. Total BS earned	0.074	0.060
28	Noble & Robinson, 1993	Post-secondary Achievement	older age peers	4. Total Masters earned	0.121	0.065
29	Noble & Robinson, 1993	Post-secondary Achievement	older age peers	5. vocational attainment	-0.014	0.065
30	Swiatek & Benbow, 1991	Post-secondary Achievement	same age peers	1. attended college	-0.227	0.541

No.	Study	Subgroup within study	Comparison group	Outcome variable	Hedges's *g*	Variance
31	Swiatek & Benbow, 1991	Post-secondary Achievement	same age peers	2. earned honors and awards in college	0.305	0.070
32	Swiatek & Benbow, 1991	Post-secondary Achievement	same age peers	3. received math/science awards up to and including college	0.032	0.078
33	Swiatek & Benbow, 1991	Post-secondary Achievement	same age peers	4. received non-math/science awards up to and including college	0.101	0.081
34	Swiatek & Benbow, 1991	Post-secondary Achievement	same age peers	5. entered college early	1.137	0.177
35	Swiatek & Benbow, 1991	Post-secondary Achievement	same age peers	6. attended graduate school	0.861	0.088
36	Swiatek & Benbow, 1991	Post-secondary Achievement	same age peers	7. created an original invention/process	0.200	0.095
37	Swiatek & Benbow, 1991	Post-secondary Achievement	same age peers	8. Authored published material	0.332	0.088
38	Gagné & Gagnier, 2004	P-12 Achievement	older age peers	Academic achievement	0.144	0.029
39	Gagné & Gagnier, 2004	P-12 Achievement	older age peers	Academic achievement	0.457	0.023

No.	Study	Subgroup within study	Comparison group	Outcome variable	Hedges's *g*	Variance
40	Mills, Ablard, & Gustin, 1994	P-12 Achievement	same age peers	1. Arithmetic/Prealgebra	0.567	0.007
41	Janos, Robinson, & Lunneborg, 1989	Post-secondary Achievement	same age peers	1. Concept Mastery Test	0.168	0.047
42	Janos, Robinson, & Lunneborg, 1989	Post-secondary Achievement	older age peers	1. Concept Matery Test	-0.452	0.047
43	Janos, Robinson, & Lunneborg, 1989	Post-secondary Achievement	older age peers	2. Total UW credits	-0.028	0.049
44	Janos, Robinson, & Lunneborg, 1989	Post-secondary Achievement	older age peers	3. Cumulative UW GPA	-0.130	0.049
45	Janos, Robinson, & Lunneborg, 1989	Post-secondary Achievement	older age peers	4. Departmental Honor Credits	-0.423	0.050
46	Janos, Robinson, & Lunneborg, 1989	Post-secondary Achievement	older age peers	5. College of Arts and Science Honor Credits	-0.718	0.052
47	Ysseldyke, Tardrew, Betts, Thill, & Hannigan, 2004	P-12 Achievement	same age peers	1. Math skill test	0.442	0.040
48	Swiatek, & Benbow, 1991 (2)	P-12 Achievement	older age peers	1. Educational level	0.538	0.019
49	Swiatek, & Benbow, 1991 (2)	P-12 Achievement	older age peers	2. Educational aspirations	0.185	0.019

No.	Study	Subgroup within study	Comparison group	Outcome variable	Hedges's *g*	Variance
50	Swiatek, & Benbow, 1991 (2)	P-12 Achievement	older age peers	3. Undergradaute GPA	-0.021	0.019
51	Swiatek, & Benbow, 1991 (2)	P-12 Achievement	older age peers	4. Number of undergraduate math course taken	0.091	0.019
52	Swiatek, & Benbow, 1991 (2)	P-12 Achievement	older age peers	5. Number of undergraduate physical science courses taken	0.152	0.019
53	Swiatek, & Benbow, 1991 (2)	P-12 Achievement	older age peers	6. Number of undergradaute natural science courses taken	-0.078	0.019
54	Swiatek, & Benbow, 1991 (2)	P-12 Achievement	older age peers	7. Number of undergraduate computer science courses taken	0.185	0.019
55	Swiatek, & Benbow, 1991 (2)	P-12 Achievement	older age peers	8. Number of unrequired mathematics course taken	0.381	0.019
56	Swiatek, & Benbow, 1991 (2)	P-12 Achievement	older age peers	9. Number of unrequired science courses taken	0.214	0.019
57	Swiatek, & Benbow, 1991 (2)	P-12 Achievement	older age peers	1. attending college	0.386	0.429
58	Swiatek, & Benbow, 1991 (2)	P-12 Achievement	older age peers	2. majoring in mathematics/science as undergraduate	-0.046	0.024

No.	Study	Subgroup within study	Comparison group	Outcome variable	Hedges's *g*	Variance
59	Swiatek, & Benbow, 1991 (2)	P-12 Achievement	older age peers	3. earning honors as undergradaute	0.000	0.028
60	Swiatek, & Benbow, 1991 (2)	P-12 Achievement	older age peers	4. attending graduate school	0.311	0.027
61	Swiatek, & Benbow, 1991 (2)	P-12 Achievement	older age peers	5. majoring in mathematics/science as graduate student	-0.022	0.023
6	Swiatek, & Benbow, 1991 (2)	P-12 Achievement	older age peers	6. creating original invention or process	0.297	0.063
63	Swiatek, & Benbow, 1991 (2)	P-12 Achievement	older age peers	7. editing a publication	0.032	0.032
64	Swiatek, & Benbow, 1991 (2)	P-12 Achievement	older age peers	8. presenting a paper; participating in a colloquium	0.347	0.031
65	Swiatek, & Benbow, 1991 (2)	P-12 Achievement	older age peers	9. publishing a book	0.000	0.570
66	Swiatek, & Benbow, 1991 (2)	P-12 Achievement	older age peers	10. publishing a journal article	0.283	0.043
67	Swiatek, & Benbow, 1991 (2)	P-12 Achievement	older age peers	11.publishing a magazine article	0.475	0.108
68	Swiatek, & Benbow, 1991 (2)	P-12 Achievement	older age peers	12.publishing a newspaper article	0.214	0.075

No.	Study	Subgroup within study	Comparison group	Outcome variable	Hedges's *g*	Variance
69	Swiatek, & Benbow, 1991 (2)	P-12 Achievement	older age peers	13. having probable publications in preparation	0.469	0.032
70	Swiatek, & Benbow, 1991 (2)	P-12 Achievement	older age peers	14. contributing to a research project	0.422	0.025
71	Swiatek, & Benbow, 1991 (2)	P-12 Achievement	older age peers	1. Oral Reading Fluency	0.137	0.026
72	Reis, Eckert, McCoach, Jacobs, & Coyne, 2008	P-12 Achievement	same age peers	2. Elementary Reading Attitude Survey	-4.145	0.030
73	Reis, Eckert, McCoach, Jacobs, & Coyne, 2008	P-12 Achievement	same age peers	3. Iowa Tests of Basic Skills	-0.017	0.010
74	Reis, Westberg, Kulikowich, & Purcell, 1998	P-12 Achievement	same age peers	1. Read	0.307	0.023
75	Reis, Westberg, Kulikowich, & Purcell, 1998	P-12 Achievement	same age peers	2. Spell	0.313	0.023
76	Reis, Westberg, Kulikowich, & Purcell, 1998	P-12 Achievement	same age peers	3. Math Concepts	0.347	0.023

No.	Study	Subgroup within study	Comparison group	Outcome variable	Hedges's *g*	Variance
77	Reis, Westberg, Kulikowich, & Purcell, 1998	P-12 Achievement	same age peers	4. Math Computation	0.454	0.023
78	Reis, Westberg, Kulikowich, & Purcell, 1998	P-12 Achievement	same age peers	5. Social Science	0.292	0.023
79	Reis, Westberg, Kulikowich, & Purcell, 1998	P-12 Achievement	same age peers	6. Science	0.406	0.023
80	Reis, Westberg, Kulikowich, & Purcell, 1998	P-12 Achievement	same age peers	1. Mathematics achievement	0.349	0.023
81	Ma, 2005	P-12 Achievement	mixed-age peers	2. Basic Skills	0.322	0.015
82	Ma, 2005	P-12 Achievement	mixed-age peers	3. Algebra	0.279	0.015
83	Ma, 2005	P-12 Achievement	mixed-age peers	4. Geometry	-0.998	0.016
84	Ma, 2005	P-12 Achievement	mixed-age peers	1. Vocabulary	-0.117	0.014

No.	Study	Subgroup within study	Comparison group	Outcome variable	Hedges's *g*	Variance
85	Moon, & Callahan, 2001	P-12 Achievement	mixed-age peers	1. Vocabulary	0.151	0.068
86	Moon, & Callahan, 2001	P-12 Achievement	mixed-age peers	1. Vocabulary	0.048	0.062
87	Moon, & Callahan, 2001	P-12 Achievement	mixed-age peers	1. Vocabulary	-0.019	0.053
88	Moon, & Callahan, 2001	P-12 Achievement	mixed-age peers	2.Reading	-0.122	0.037
89	Moon, & Callahan, 2001	P-12 Achievement	mixed-age peers	2.Reading	0.058	0.068
90	Moon, & Callahan, 2001	P-12 Achievement	mixed-age peers	2.Reading	0.263	0.062
91	Moon, & Callahan, 2001	P-12 Achievement	mixed-age peers	2.Reading	0.112	0.053
92	Moon, & Callahan, 2001	P-12 Achievement	mixed-age peers	3. Language	-0.133	0.037
93	Moon, & Callahan, 2001	P-12 Achievement	mixed-age peers	3. Language	-0.105	0.068
94	Moon, & Callahan, 2001	P-12 Achievement	mixed-age peers	3. Language	0.335	0.063

No.	Study	Subgroup within study	Comparison group	Outcome variable	Hedges's *g*	Variance
95	Moon, & Callahan, 2001	P-12 Achievement	mixed-age peers	3. Language	0.080	0.053
96	Moon, & Callahan, 2001	P-12 Achievement	mixed-age peers	4. Mathematics	-0.329	0.038
97	Moon, & Callahan, 2001	P-12 Achievement	mixed-age peers	4. Mathematics	0.137	0.068
98	Moon, & Callahan, 2001	P-12 Achievement	mixed-age peers	4. Mathematics	0.032	0.062
99	Moon, & Callahan, 2001	P-12 Achievement	mixed-age peers	4. Mathematics	-0.031	0.053
100	Moon, & Callahan, 2001	P-12 Achievement	mixed-age peers	1. High school grades---achieved A+	0.037	0.053
101	Barnett & Durden, 1993	P-12 Achievement	mixed-age peers	2. College-level courses in high school	0.329	0.014
102	Barnett & Durden, 1993	P-12 Achievement	mixed-age peers	3. Tests of achievement in specific disciplines	0.458	0.013
103	Barnett & Durden, 1993	Post-secondary Achievement	mixed-age peers	1. Selectivity of college attended	0.294	0.013
104	Brody, 1985	P-12 Achievement	same age peers	1. SAT-Verbal score---language	-0.097	0.049

No.	Study	Subgroup within study	Comparison group	Outcome variable	Hedges's *g*	Variance
105	Brody, 1985	P-12 Achievement	same age peers	1. SAT-Verbal score---writing	-0.029	0.030
106	Brody, 1985	P-12 Achievement	same age peers	1. SAT-Verbal score---science	0.251	0.056
107	Brody, 1985	P-12 Achievement	same age peers	1. SAT-Verbal score-math	0.145	0.051
108	Brody, 1985	P-12 Achievement	same age peers	1. SAT-Verbal score---Total average	0.058	0.026
109	Brody, 1985	P-12 Achievement	same age peers	2. SAT-Math score---language	-0.152	0.049
110	Brody, 1985	P-12 Achievement	same age peers	2. SAT-Math score---writing	-0.181	0.030
111	Brody, 1985	P-12 Achievement	same age peers	2. SAT-Math score---science	0.164	0.056
112	Brody, 1985	P-12 Achievement	same age peers	2. SAT-Math score---math	0.228	0.051
113	Brody, 1985	P-12 Achievement	same age peers	2. SAT-Math score---Total Average	-0.002	0.026
114	Weiner, 1985	Post-secondary Achievement	same age peers	1.Academic	0.078	0.033

No.	Study	Subgroup within study	Comparison group	Outcome variable	Hedges's *g*	Variance
115	Weiner, 1985	Post-secondary Achievement	same age peers	2.Creativity	0.280	0.033
116	Weiner, 1985	Post-secondary Achievement	same age peers	3.Artistic	0.231	0.033
117	Weiner, 1985	Post-secondary Achievement	same age peers	4.Leadership	0.374	0.033
118	Weiner, 1985	Post-secondary Achievement	same age peers	5.Vocational	0.540	0.034
119	Weiner, 1985	Post-secondary Achievement	same age peers	6.Educational	0.438	0.034
120	Weiner, 1985	Post-secondary Achievement	same age peers	Cooperative Mathematics Test, Algebra II	0.325	0.033
121	Washington, 1999	P-12 Achievement	same age peers	Cooperative Mathematics Test, Algebra II	0.462	0.016
122	Washington, 1999	P-12 Achievement	same age peers	Cooperative Mathematics Test, Algebra II	0.822	0.007
123	Washington, 1999	P-12 Achievement	same age peers	Cooperative Mathematics Test, Algebra II	0.815	0.068
124	Washington, 1999	P-12 Achievement	same age peers	Cooperative Mathematics Test, Algebra II	0.328	0.022

No.	Study	Subgroup within study	Comparison group	Outcome variable	Hedges's *g*	Variance
125	Washington, 1999	P-12 Achievement	same age peers	Cooperative Mathematics Test, Algebra II	0.657	0.012
126	Ambruster, 1995	P-12 Achievement	mixed-age peers	Cummulative Grade Point Average	-0.067	0.013
127	Ambruster, 1995	P-12 Achievement	mixed-age peers	Cummulative Grade Point Average	-0.032	0.015
128	Ambruster, 1995	P-12 Achievement	mixed-age peers	Cummulative Grade Point Average	-0.030	0.021
129	Ambruster, 1995	P-12 Achievement	mixed-age peers	Cummulative Grade Point Average	-0.005	0.027
130	Ambruster, 1995	P-12 Achievement	mixed-age peers	Cummulative Grade Point Average	0.056	0.039
131	Ambruster, 1995	P-12 Achievement	mixed-age peers	Cummulative Grade Point Average	-0.076	0.073
132	Ambruster, 1995	P-12 Achievement	mixed-age peers	Cummulative Grade Point Average	0.000	0.023
133	Janos, 1987	Post-secondary Achievement	same age peers	The mean age when finished advanced degrees	-1.044	0.312
134	Richards, 2006	Post-secondary Achievement	same age peers	College course grades	0.099	0.002

No.	Study	Subgroup within study	Comparison group	Outcome variable	Hedges's *g*	Variance
135	Fowler, 2007	P-12 Achievement	same age peers	retention of students in high school	0.231	0.008
136	Fowler, 2007	P-12 Achievement	same age peers	high school graduation rates	0.175	0.008
137	Fowler, 2007	P-12 Achievement	same age peers	post-secondary school enrollment	0.211	0.008
138	Chilton, 2001	P-12 Achievement	same age peers	Standardized achievement test in high school	0.804	0.028
139	Chilton, 2001	P-12 Achievement	same age peers	Highest math class taken	-0.053	0.067
140	Chilton, 2001	P-12 Achievement	same age peers	Math class enrollment	0.000	0.063
141	Wells, Lohman, & Marron, 2008	P-12 Achievement	older age peers	Academic and acceleration outcomes (NELS data-1992)	0.685	0.052
142	Wells, Lohman, & Marron, 2008	P-12 Achievement	older age peers	Academic and acceleration outcomes (ELS data-2004)	0.618	0.085
143	Reynolds, 1993	P-12 Achievement	older age peers	academic grades	0.272	0.058
144	Reynolds, 1993	P-12 Achievement	older age peers	Total reading score (Stanford-8 Achievement Test)	0.375	0.063

No.	Study	Subgroup within study	Comparison group	Outcome variable	Hedges's *g*	Variance
145	Reynolds, 1993	P-12 Achievement	older age peers	Total math score (Stanford-8 Achievement Test)	0.138	0.062
146	Reynolds, 1993	P-12 Achievement	older age peers	Total language score (Stanford-8 Achievement Test)	0.344	0.064
147	Reynolds, 1993	P-12 Achievement	older age peers	Otis Lennon School Abilities Test	0.620	0.065

Appendix E. Social-emotional Development Study Outcomes and Effect Sizes Extracted

No.	Study	Comparison group	Outcome variable	Hedges's *g*	Variance
1	Janos & Robinson, 1985	older age peers	1. perception of importance of academic characteristics	0.490	0.096
2	Janos & Robinson, 1985	older age peers	2. Perception of satisfaction with academic characteristics	0.658	0.098
3	Brody & Benbow, 1987	mixed-age peers	1. No. of Types of activities	-0.061	0.018
4	Brody & Benbow, 1987	mixed-age peers	2. No. of Leadership activities	-0.200	0.018
5	Brody & Benbow, 1987	mixed-age peers	3. Self-esteem	0.133	0.018
6	Brody & Benbow, 1987	mixed-age peers	4. Locus of control	0.037	0.018
7	Cornell, Gallahan, & Loyd, 1991	same age peers	1. personality growth	-0.015	0.083
8	Cornell, Gallahan, & Loyd, 1991	same age peers	1. personality growth	0.234	0.084
9	Sayley & Brookshire, 1993	older age peers	1. locus of control	0.054	0.006
10	Sayley & Brookshire, 1993	older age peers	2. Self-concept	-0.029	0.006
11	Sayley & Brookshire, 1993	older age peers	3. Peer perceptions	-0.056	0.008
12	Sayley & Brookshire, 1993	older age peers	4. Behavioral problems	0.125	0.008
13	Noble & Robinson, 1993	older age peers	1. like about undergraduate experience	0.066	0.067
14	Noble & Robinson, 1993	older age peers	2. see themselves in a full-time career plan	0.335	0.072
15	Noble & Robinson, 1993	older age peers	3. see themsevels in a part-time career plan	0.114	0.158

No.	Study	Comparison group	Outcome variable	Hedges's *g*	Variance
16	Noble & Robinson, 1993	older age peers	4. participate activities during undergraduate	-0.210	0.048
17	Noble & Robinson, 1993	older age peers	5. Positive beliefs and attitudes	-0.157	0.047
18	Noble & Robinson, 1993	older age peers	6. Positive values	0.127	0.053
19	Noble & Robinson, 1993	same age peers	6. Positive values	0.048	0.043
20	Stamps, 2004	same age peers	Motivation-Students attitudes toward course content	0.168	0.057
21	Robinson & Janos, 1986	same age peers	1. the Minnesota Multiphasic Personality Inventory	-0.023	0.076
22	Robinson & Janos, 1986	same age peers	2.the California Psychological Inventory	0.043	0.076
23	Robinson & Janos, 1986	older age peers	1. the Minnesota Multiphasic Personality Inventory	0.072	0.082
24	Robinson & Janos, 1986	older age peers	2.the California Psychological Inventory	-0.239	0.083
25	Swiatek & Benbow, 1991	same age peers	1. self-esteem	-0.528	0.045
26	Gagné & Gagnier, 2004	older age peers	1. Conduct	-0.043	0.029
27	Gagné & Gagnier, 2004	older age peers	1. Conduct	0.289	0.022
28	Gagné & Gagnier, 2004	older age peers	2. Integration	0.192	0.029
29	Gagné & Gagnier, 2004	older age peers	2. Integration	0.249	0.022
30	Gagné & Gagnier, 2004	older age peers	3. Maturity	0.104	0.029

No.	Study	Comparison group	Outcome variable	Hedges's *g*	Variance
31	Gagné & Gagnier, 2004	older age peers	3. Maturity	0.251	0.022
32	Sayler, 2008	older age peers	1. Standard of living	-0.164	0.047
33	Sayler, 2008	older age peers	2. Health	-0.413	0.048
34	Sayler, 2008	older age peers	3. Achieving in life	-0.399	0.048
35	Sayler, 2008	older age peers	4. Personal relationships	-0.428	0.048
36	Sayler, 2008	older age peers	5. Safety	-0.314	0.047
37	Sayler, 2008	older age peers	6. Community	-0.489	0.048
38	Sayler, 2008	older age peers	7. Future security	0.185	0.047
39	Sayler, 2008	older age peers	8. Composite score	-0.428	0.048
40	Janos, Robinson, & Lunneborg, 1989	same age peers	1. Dominance	-0.048	0.045
41	Janos, Robinson, & Lunneborg, 1989	same age peers	2. Capacity for Status	-0.019	0.045
42	Janos, Robinson, & Lunneborg, 1989	same age peers	3. Sociability	0.146	0.045
43	Janos, Robinson, & Lunneborg, 1989	same age peers	4. Social presence	0.026	0.045
44	Janos, Robinson, & Lunneborg, 1989	same age peers	5. Self-acceptance	-0.105	0.045
45	Janos, Robinson, & Lunneborg, 1989	same age peers	6. Well Being	0.106	0.045

No.	Study	Comparison group	Outcome variable	Hedges's *g*	Variance
46	Janos, Robinson, & Lunneborg, 1989	same age peers	7. Responsibility	-0.119	0.045
47	Janos, Robinson, & Lunneborg, 1989	same age peers	8. Socialization	0.047	0.045
48	Janos, Robinson, & Lunneborg, 1989	same age peers	9.Self-control	0.253	0.046
49	Janos, Robinson, & Lunneborg, 1989	same age peers	10. Tolerance	0.372	0.046
50	Janos, Robinson, & Lunneborg, 1989	same age peers	11. Good Impression	0.190	0.045
51	Janos, Robinson, & Lunneborg, 1989	same age peers	12. Communality	0.140	0.045
52	Janos, Robinson, & Lunneborg, 1989	same age peers	13. Achievement via conformance	0.249	0.046
53	Janos, Robinson, & Lunneborg, 1989	same age peers	14. Achievement via independence	0.320	0.046
54	Janos, Robinson, & Lunneborg, 1989	same age peers	15. Intellectual efficiency	0.099	0.045
55	Janos, Robinson, & Lunneborg, 1989	same age peers	16. Psychological mindedness	0.274	0.046

No.	Study	Comparison group	Outcome variable	Hedges's *g*	Variance
56	Janos, Robinson, & Lunneborg, 1989	same age peers	17. Flexibility	0.258	0.046
57	Janos, Robinson, & Lunneborg, 1989	same age peers	18. Feminity	0.314	0.046
58	Janos, Robinson, & Lunneborg, 1989	same age peers	1. Social Maturity	0.238	0.045
59	Janos, Robinson, & Lunneborg, 1989	older age peers	1. Dominance	-0.182	0.046
60	Janos, Robinson, & Lunneborg, 1989	older age peers	2. Capacity for Status	0.071	0.046
61	Janos, Robinson, & Lunneborg, 1989	older age peers	3. Sociability	0.041	0.046
62	Janos, Robinson, & Lunneborg, 1989	older age peers	4. Social presence	0.083	0.046
63	Janos, Robinson, & Lunneborg, 1989	older age peers	5. Self-acceptance	-0.109	0.046
64	Janos, Robinson, & Lunneborg, 1989	older age peers	6. Well Being	-0.401	0.047
65	Janos, Robinson, & Lunneborg, 1989	older age peers	7. Responsibility	-0.250	0.047
66	Janos, Robinson, & Lunneborg, 1989	older age peers	8. Socialization	-0.300	0.047

No.	Study	Comparison group	Outcome variable	Hedges's *g*	Variance
67	Janos, Robinson, & Lunneborg, 1989	older age peers	9.Self-control	-0.189	0.046
68	Janos, Robinson, & Lunneborg, 1989	older age peers	10. Tolerance	-0.044	0.046
69	Janos, Robinson, & Lunneborg, 1989	older age peers	11. Good Impression	0.194	0.046
70	Janos, Robinson, & Lunneborg, 1989	older age peers	12. Communality	-0.247	0.047
71	Janos, Robinson, & Lunneborg, 1989	older age peers	13. Achievement via conformance	-0.284	0.047
72	Janos, Robinson, & Lunneborg, 1989	older age peers	14. Achievement via independence	-0.219	0.046
73	Janos, Robinson, & Lunneborg, 1989	older age peers	15. Intellectual efficiency	-0.443	0.047
74	Janos, Robinson, & Lunneborg, 1989	older age peers	16. Psychological mindedness	-0.116	0.046
75	Janos, Robinson, & Lunneborg, 1989	older age peers	17. Flexibility	0.315	0.047
76	Janos, Robinson, & Lunneborg, 1989	older age peers	18. Feminity	0.156	0.046

No.	Study	Comparison group	Outcome variable	Hedges's *g*	Variance
77	Janos, Robinson, & Lunneborg, 1989	older age peers	1. Social Maturity	-0.461	0.047
78	Jin & Moon, 2006	same age peers	1. Psychological Well-Being	0.047	0.016
79	Jin & Moon, 2006	same age peers	2. Satisfaction With School Life	1.281	0.017
80	Swiatek, & Benbow, 1991 (2)	older age peers	1. Liking for college	-0.107	0.019
81	Swiatek, & Benbow, 1991 (2)	older age peers	2. Number of extracurricular activity areas	0.044	0.019
82	Swiatek, & Benbow, 1991 (2)	older age peers	3. Confidence in mathematics	0.106	0.019
83	Swiatek, & Benbow, 1991 (2)	older age peers	4. Confidence in science	0.173	0.019
84	Swiatek, & Benbow, 1991 (2)	older age peers	5. Perceived ease of mathematics	0.055	0.019
85	Swiatek, & Benbow, 1991 (2)	older age peers	6. Perceived ease of science	0.019	0.019
86	Swiatek, & Benbow, 1991 (2)	older age peers	7. Interest in mathematics	0.154	0.019
87	Swiatek, & Benbow, 1991 (2)	older age peers	8. Interest in science	0.043	0.019
88	Swiatek, & Benbow, 1991 (2)	older age peers	9. Usefulness of mathematics for planned career	0.088	0.019
89	Swiatek, & Benbow, 1991 (2)	older age peers	10. Usefulness of science for planned career	0.054	0.019
90	Swiatek, & Benbow, 1991 (2)	older age peers	11. Attitudes toward math scale	0.156	0.019
91	Swiatek, & Benbow, 1991 (2)	older age peers	12. Attitudes toward science scale	0.110	0.019
92	Swiatek, & Benbow, 1991 (2)	older age peers	13. Locus of control scale	-0.213	0.019
93	Swiatek, & Benbow, 1991 (2)	older age peers	14. Self-esteem scale	0.049	0.019

No.	Study	Comparison group	Outcome variable	Hedges's *g*	Variance
94	Plucker, & Taylor, 1998	same age peers	1. Self-concept	0.041	0.011
95	Ma, 2002	mixed-age peers	1. self-esteem	0.063	0.014
96	Washington, 1999	same age peers	1. self-concept	0.100	0.013
97	Washington, 1999	same age peers	1. self-concept	0.094	0.016
98	Washington, 1999	same age peers	1. self-concept	0.252	0.006
99	Washington, 1999	same age peers	1. self-concept	0.397	0.064
100	Washington, 1999	same age peers	1. self-concept	-0.003	0.021
101	Washington, 1999	same age peers	1. self-concept	0.117	0.012
102	Ambruster, 1995	mixed-age peers	Social-Behavioral	-0.056	0.013
103	Ambruster, 1995	mixed-age peers	Social-Behavioral	-0.469	0.016
104	Ambruster, 1995	mixed-age peers	Social-Behavioral	-0.535	0.021
105	Ambruster, 1995	mixed-age peers	Social-Behavioral	-0.255	0.027
106	Ambruster, 1995	mixed-age peers	Social-Behavioral	-0.073	0.040
107	Ambruster, 1995	mixed-age peers	Social-Behavioral	0.293	0.074
108	Ambruster, 1995	mixed-age peers	Social-Behavioral	-0.254	0.023
109	Chilton, 2001	same age peers	1. Extra curricular activities	0.154	0.045
110	Chilton, 2001	same age peers	2. Leadership positions and awards	0.864	0.043
111	Chilton, 2001	same age peers	3. Student Social Experiences	0.019	0.029
112	Chilton, 2001	same age peers	4. College expectations	0.182	0.042
113	Chilton, 2001	same age peers	5. Student attitudes towards math curriculum	-0.109	0.029

No.	Study	Comparison group	Outcome variable	Hedges's *g*	Variance
114	Chilton, 2001	same age peers	6. Student beliefs regarding math ability	0.010	0.029
115	Chilton, 2001	same age peers	7. Student beliefs regarding math achievement	0.039	0.030
116	Chilton, 2001	same age peers	8. Satisfaction with Math education	-0.176	0.029
117	Chilton, 2001	same age peers	9. Enjoyment in math classes	-0.041	0.029
118	Lupkowski, Whitmore, & Ramsay, 1992	same age peers	Self-esteem (the Coppersmith Self-esteem Inventory)	0.163	0.005
119	Ingersoll, 1992	older age peers	Social skill	0.269	0.094
120	Ingersoll, 1992	older age peers	Social support	-0.746	0.099
121	Ingersoll, 1992	older age peers	Social activities	-0.621	0.132
122	Ingersoll, 1992	older age peers	CPI socialization	0.267	0.099
123	Ingersoll, 1992	older age peers	CPI Social Confidence Scale	-0.641	0.103
124	Ingersoll, 1992	older age peers	Friendship scale	0.446	0.109
125	Houston, 1999	same age peers	1. education expectations	-0.015	0.098
126	Houston, 1999	same age peers	2. Career aspirations	0.180	0.099
127	Reynolds, 1993	older age peers	1. conduct grades	0.440	0.059
128	Reynolds, 1993	older age peers	2. Extra curricular activities	0.720	0.061
129	Reynolds, 1993	older age peers	3. Out-of-school extra curricular activities	0.535	0.060
130	Reynolds, 1993	older age peers	4. Social adjustment	0.051	0.058

VITA

VITA

EDUCATION

Ph. D. Purdue University (August, 2009)

Major: Educational Psychology, specializing in gifted/talented education. Advisor: Sidney Moon

M.A. University of Science & Technology of China, 2000

Major: Curriculum and Instruction

B.A. Xiangtan Normal University, 1997

Major: Curriculum and Instruction (Political Education)

RESEARCH IN PROGRESS

Dissertation: *A Meta-Analysis of the Effects of Acceleration on High-Ability Learners*. Committee: Sidney Moon (chair), Marcia Gentry, Rebecca Mann, Yan Ping Xin

Dai, D. Y., Fu, M., & Hu, S. (in progress). Comparing accelerated and non-accelerated gifted development at secondary and college levels: Chinese experiences. Research grants ($25,000) from the Institute for Research and Policy on Acceleration, University of Iowa.

Hu, S. & Moon, S. (in progress). Using Curriculum Vitae as Research Data: Extant Practices, Patterns and Suggestions for Future Research. Proposal accepted to present at the 2008 annual meeting of American Educational Research Association, New York City, NY.

Gentry, M., & Hu, S. (under review). Student-identified exemplary teachers: What makes them tick? *The Journal of Experimental Education*.

PUBLICATIONS

Peer-Reviewed Articles

Gentry M., Hu, S., Peters, S., & Rizza, M.G (2008). Gifted students in an exemplary career and technical education school: A qualitative inquiry. *Gifted Child Quarterly, 52*, 183-198.

Gentry, M., Rizza, M.G., Peters, S., & Hu, S. (2005). Professionalism, sense of community and reason to learn: Lessons from an exemplary career and technical education center. *Career and Technical Education Research, 30*(1), 25-85.

Hu, S. & Yang, Y. (2002). Educational tenet reorientation of the Special Class for the Gifted Young in the progress of mass higher education. *The seminar collection of thesis of special education of both China and Taiwan in 2002,* 2002 (12).

Hu, S. & Sun, X. (1999). Duxiu Chen's instructions to the youth in the May 4th Movement Period. *Anhui Journal of History Research, 1999(2)* (embodied by the *Reprography Datum Center of Renming University of China*)

Dong, Y., Chang, J., & Hu, S. (1999). Impacts of 'San Xia Xiang' Program on college students. *Education and Modernization, 1999(2).*

Chang, J. & Hu, S. (1999). Featured as a complete system: Deng Xiaoping's theory of socialism construction. *Debate Terrace of Hefei*, *1999(1).*

Sun, X. & Hu, S. (1999). Analysis of Deng Xiaoping's thoughts regarding diplomatic strategy to Japan. *The Great Change-----Memory for the Central Committee's 3rd Total Session of the 11th Congress of the CPC*.

Hu, S. & Xun, X. (1998). Exploration of Deng Xiaoping's strategic thoughts on Japan. *Theory and Practice of Mao Zedong Thoughts, 1998*(*4).*

Book Chapters

Gentry, M., Hu, S., & Thomas, A. (2007). Ethnically diverse gifted students: Research findings, implications, and future directions. In C. M. Callahan & J. A. Plucker (Eds.), *Critical issues and practices in gifted education: What research says* (pp.195-212). Waco, Texas: Prufrock Press.

Moon, S., & Hu, S. (2007). Personal talent research. In C. M. Callahan & J. A. Plucker (Eds.), *Critical issues and practices in gifted education: What research says* (pp.493-511). Waco, Texas: Prufrock Press.

PRESENTATIONS AT NATIONAL CONFERENCES

Steenbergen-Hu. S. (2008, October). *A meta-analysis of the effects of acceleration on high ability learners.* Presentation at the 2008 annual meeting of National Association for Gifted Children, Tampa, FL.

Yoon, S. Y., Hu. S., & Gentry, M. (2007, November). The role of Engineering in the K-12 Education: Implications for Gifted Education. Proposal presented at the 2007 annual meeting of National Association for Gifted Children, Minneapolis Twin-Cities, MN.

Hu. S. (2007, November). *Does acceleration serve gifted learners well? A meta-analysis of the effects of acceleration.* Presentation at the 2007 annual meeting of National Association for Gifted Children, Minneapolis Twin-Cities, MN.[Graduate Student Research Gala].

Gentry, M. & Hu. S. (2006, November). Gifted teachers: What makes them tick? Presentation at the annual meeting of National Association for Gifted Children, Charlotte, NC.

Yoon, S. Y. & Hu, S. (2006, November). Overrepresentation phenomena of Asian American students in gifted programs: A Search for statistical evidence and explanations. Presentation at the annual meeting of National Association for Gifted Children, Charlotte, NC. [Graduate Student Research Gala].

Yoon, S.Y. & Hu, S. (2006, May). Gifted East Asian Americans: A close looking at the overrepresentation issue. Poster session presented at The Eighth Biennial Henry B. & Jocelyn Wallace National Research Symposium on Talent Development, Iowa City, IA.

Gentry M., Hu, S., & Peters, S. (2006, April). Gifted students in an exemplary career and technical education school: A qualitative inquiry. Presentation at the annual meeting of American Educational Research Association, San Francisco, CA.

Gentry, M., Choi, B.Y., & Hu, S. (2005, November). Exemplary teachers: Students' perspectives and teachers' practices. Presentation at the annual meeting of National Association for Gifted Children, Louisville, Kentucky.

TECHNICAL REPORTS

Hu, S. (2007). Evaluation report of 2007 *Summer Residential and Super Summer* programs. Purdue University, West Lafayette, IN.

Hu, S., & Moon, S. (2005). *Science bound summer camp evaluation report*. Purdue University, West Lafayette, IN.

Hu, S. (2005). Evaluation report of 2005 *Summer Residential and Super Summer* programs. Purdue University, West Lafayette, IN.

PROFESSIONAL EXPERIENCE

Research Experience

Research assistant, Purdue University, College of Education, August 2004-2008

- Literature searches
- Data analysis
- Research synthesizing
- Manuscript preparation for publication

Program coordinator, worked on evaluations of student programs

Gifted Education Resources Institute, Purdue University, August 2004-Present

- Survey designing
- Survey/data collection
- Data entry and analysis
- Report writing

Teaching Experience

Instructor, University of Science & Technology of China, June 2000-August 2004

Courses taught:

(1) *Impacts of Famous Scientists on College Students*

(2) *Political Education*

- Designed and taught 20 weeks undergraduate course for 7 semesters
- Seven hours average teaching hours per week
- Teaching evaluations ranked top 10% for three times

Class supervisor, Special Class for the Gifted Young (Dept. of Gifted Education)

- Mentored and consulted 105 undergraduates
- Directed extracurricular activities
- Evaluated student achievement

Assistant to the Dean, Special Class for the Gifted Young (Dept. of Gifted Education), University of Science & Technology of China, June 2000-August 2004

- Assisted the design of curriculum and instruction
- Drafted documents for the administration
- Corresponded with alumni and visiting scholars and journalists

RESEARCH &TEACHING INTERESTS

Research Interests

- Research synthesis and meta-analysis
- Motivation and productivity
- Talent development and psychology
- Decision-making

Teaching Interests

- Research methodology
- Research synthesis methodology and practices
- Learning and instruction strategies and practice

HONORS AND AWARDS

2008	A research paper was recognized as the First Place among doctoral research in progress at the Graduate Student Research Gala, at the 2008 annual meeting of National Association for Gifted Children, Tampa, FL.
2007	First recipient of the John Feldhusen Fellowship College of Education, Purdue University
2004- 2008	Ross Doctoral Fellowship, Dept. of Educational Studies, Purdue University
2002	Huawei Teaching Fellowship University of Science & Technology of China
2001	Outstanding Teacher and Huawei Teacher Awards University of Science & Technology of China
1998	Outstanding Student Leader University of Science & Technology of China
1998	Di'Ao Scholarship University of Science & Technology of China

PROFESSIONAL AFFILIATIONS

2007-Present	Toastmaster International (public speaking club)
2006-2007	American Educational Research Association
2005-present	National Association for Gifted Children
2002-2004	Council for the Gifted and Talented of Chinese Research Association