Review of Educational Research
December 2016, Vol. 86, No. 4, pp. 849–899
DOI: 10.3102/0034654316675417

What One Hundred Years of Research Says About the Effects of Ability Grouping and Acceleration on K–12 Students' Academic Achievement: Findings of Two Second-Order Meta-Analyses

Saiying Steenbergen-Hu
Northwestern University

Matthew C. Makel
Duke University

Paula Olszewski-Kubilius
Northwestern University

Two second-order meta-analyses synthesized approximately 100 years of research on the effects of ability grouping and acceleration on K–12 students' academic achievement. Outcomes of 13 ability grouping meta-analyses showed that students benefited from within-class grouping ($0.19 \leq g \leq 0.30$), cross-grade subject grouping ($g = 0.26$), and special grouping for the gifted ($g = 0.37$), but did not benefit from between-class grouping ($0.04 \leq g \leq 0.06$); the effects did not vary for high-, medium-, and low-ability students. Three acceleration meta-analyses showed that accelerated students significantly outperformed their nonaccelerated same-age peers ($g = 0.70$) but did not differ significantly from nonaccelerated older peers ($g = 0.09$). Three other meta-analyses that aggregated outcomes across specific forms of acceleration found that acceleration appeared to have a positive, moderate, and statistically significant impact on students' academic achievement ($g = 0.42$).

Keywords: ability grouping, acceleration, effect size, meta-analysis, second-order meta-analysis

According to the U.S. Census Bureau, in the fiscal year 2013, total expenditures at all governmental levels (including capital outlays) on public elementary and secondary schools was $596.3 billion. This amount does not include spending on public college and universities, nor the roughly 10% (U.S. Department of Education, 2014) of students who attended private schools. We mention this figure to highlight the vast resources devoted to education in the United States. Whether these vast resources are allocated to maximize the development of high-ability students' talents remains in question. A recent policy brief reported that 20% to 40% of elementary and middle school students perform above grade level in reading and 10% to 30% do so in math (Makel, Matthews, Peters, Rambo-Hernandez, & Plucker, 2016). With so many students performing above grade level, the authors concluded that the U.S. educational context requires major changes to provide such students with opportunities to learn. However, many researchers have expressed concern about the lack of empirical evidence supporting the effectiveness of special programming and interventions for gifted students, often citing studies demonstrating no effect or even potential harm (Adelson, McCoach, & Gavin, 2012; Bui, Craig, & Imberman, 2011; Hattie, 2002; Oakes, 1985; Slavin, 1987, 1990). At the same time, there is evidence supporting the efficacy of interventions and instructional strategies for students with advanced talent in academic domains (e.g., Assouline, Colangelo, VanTassel-Baska, & Lupkowski-Shoplik, 2015; Lubinski & Benbow, 2000).

When a subject has particularly strong political and policy implications and the evidence is inconsistent across individual studies, particularly when a large corpus of evidence exists, comprehensive syntheses of evidence are particularly useful. To this end, the current study consists of two second-order meta-analyses of existing meta-analytic studies that aggregated the outcomes of empirical primary studies on the effects of ability grouping and academic acceleration (acceleration for short) on K–12 students' academic achievement. Ability grouping and acceleration are educational interventions that seek to promote learning for high-achieving and high-ability students. Although these groups share similarities, researchers typically treat them as distinct. Detailed discussion of their similarities and differences are beyond the scope of the current study, but we provide definitions and relevant background information below.

Ability Grouping

Defining Ability Grouping

There are great misconceptions surrounding the term "ability grouping." As Oakes (1985) noted, ability grouping means different things to different people at different times. Many have used terms such as tracking, streaming, setting, sorting, classroom organization or composition, and classroom assignment. Although terms such as tracking and ability grouping have been used interchangeably in the past, researchers differentiate ability grouping from tracking. Although both ability grouping and tracking involve assigning students based on their prior achievement or ability levels (Loveless, 2013), the former often takes place in elementary schools with the latter occurring in middle and high schools. Other researchers, such as Tieso (2003), argue that ability grouping is a more flexible form of grouping than tracking.

In the current study, we define ability grouping as an instructional practice with three key features: (a) it involves placing students into different classrooms or small groups based on their initial achievement skill levels, readiness, or abilities;

(b) the main purpose of such placement is to create a more homogeneous learning environment so that teachers can provide instruction better matched to students' needs and so that students can benefit from interactions with their comparable academic peers; and (c) such placements are not permanent school administrative arrangements that lead to restrictions on students' graduation, destination, or career paths. With this definition, we intend to differentiate ability grouping from historical tracking systems that involved assigning students (mostly middle and high school students) to fixed academic, general, or vocational tracks based primarily on their ability, achievement levels, or career aspirations (Chmielewski, 2014; Loveless, 2009).

Ability grouping takes various forms. On the basis of our comprehensive review of the literature, we categorized ability grouping into four main types. The first is between-class ability grouping, which involves assigning students of the same grade into high, average, or low classes based on their prior achievement or ability levels. This form of ability grouping has been labeled differently in different publications, including comprehensive ability-grouped classes (Slavin, 1987, 1990, 1993), XYZ groupings (Mosteller, Light, & Sachs, 1996), between-class comprehensive grouping (Kulik & Kulik, 1987), and multilevel classes (Kulik & Kulik, 1992). The second type is within-class ability grouping—also called small-group instruction (Lou et al., 1996)—which involves teachers assigning students within a class to several small homogeneous groups for instruction based on students' prior achievement or learning capacities. This type of grouping has most frequently been used in elementary classrooms. Cluster grouping or total school cluster grouping is a type of within-class grouping because it places students identified as gifted, high-achieving, or high-ability into classrooms that consist of students of other achievement levels to affect the composition of the classroom, to facilitate learning through differentiation, and to improve student achievement (Gentry, Paul, McIntosh, Fugate, & Jen, 2014). We consider cluster grouping to be conceptually relevant to our second-order meta-analysis of ability grouping. However, cluster grouping was rarely mentioned in the 13 ability grouping meta-analyses that were eligible for the current second-order meta-analyses. Additionally, our literature search identified no meta-analysis of the effects of cluster grouping to date.

The third type of ability grouping is cross-grade subject grouping, which involves grouping students of different grade levels together to learn a particular subject based on their prior achievement or learning potential. The Joplin Plan (Floyd, 1954), which groups students of different grade levels for reading instruction, is the best-known and most representative type of cross-grade subject grouping. The last type of ability grouping is special grouping for the gifted, which often refers to educational and instructional programs that were designed specifically for gifted and talented students, such as pull-out or honors programs. It is important to note that several major meta-analyses of ability grouping (Lou et al., 1996; Mosteller et al., 1996; Slavin, 1987, 1990, 1993) only included studies of general student populations and explicitly excluded studies of grouping interventions for gifted and talented students. However, at least six meta-analyses included studies of grouping for gifted and talented students (Goldring, 1990; Kulik, 1985; Kulik & Kulik, 1982, 1984a, 1987, 1992).

The History of Ability Grouping (1960s to Present)

Ability grouping was widely embraced in U.S. school systems from the 1960s to the end of the 1980s. However, this practice began to fall out of fashion from the middle of the 1980s to the end of the 1990s, partly as a result of opposition from advocates for equity and equality, most notably, Jeannie Oakes (1985) and Robert Slavin (1987, 1990, 1993). For example, Oakes (1985) argued that tracking unfairly limited educational opportunities for disadvantaged students, thus exacerbating existing educational and social inequalities. In the same period, Slavin's (1987, 1990, 1993) best-evidence syntheses concluded that the effects of ability grouping on elementary, secondary, and middle school student achievement were essentially zero. By the mid-1990s, many schools, especially high-poverty urban middle schools, were reducing or even eliminating tracking (Loveless, 2009). Based on data from the National Center for Education Statistics, the Brookings Institution found that only 40% of teachers reported grouping students for mathematics instruction in 1996.

However, ability grouping practices have increased markedly since the end of the 1990s and have been gaining in popularity in recent years. Citing data from the National Center for Education Statistics, Loveless (2009) noted that tracking is on the rise and affects more than 14 million middle-grade students annually. As presented in Loveless's (2009, p. 17) *2013 Brown Center Report on American Education*, National Assessment of Educational Progress data showed that "the percentage of students placed into ability groups for reading instruction skyrocketed from 1998 to 2009, from 28% to 71%. . . . Math ability grouping . . . accelerates from 2003 to 2011 (reaching 61% in 2011)." It is relevant to note that teachers' beliefs about the influence of student heterogeneity on instruction seem to undergird the use of ability grouping. In response to the 2008 *MetLife Survey of the American Teacher*'s survey statement, "My class/classes in my school have become so mixed in terms of students' learning ability that I/teachers can't teach them," 14% of teacher respondents answered *agree strongly* and 29% said *agree somewhat* (Markow & Cooper, 2008).

The Ongoing Debate

Ability grouping has been one of the most controversial educational practices for more than a century. Proponents argue for its value in effectively addressing the educational needs of students whose prior achievement, skills, or abilities vary greatly (Tieso, 2003). Critics and opponents cite ability grouping as a contributor to achievement gaps, the stratification of educational opportunities, and detrimental psychosocial outcomes, such as lowered self-concept or self-esteem, particularly for disadvantaged or lower achieving students (Belfi, Goos, De Fraine, & Van Damme, 2012; Oakes, 2008). Regardless of the nature or extent of these disputes, the practical implications of ability grouping are profound. Ability grouping policies and practices affect students' experiences in school, including the courses they take, the curricula they receive, the peers with whom they learn, and the teachers who provide instruction.

Acceleration

Acceleration allows students to progress through school at a more rapid pace than their peers or to take courses at ages younger than typical students (Pressey, 1949). However, Lubinski and Benbow (2000) noted that the term *acceleration* is a misnomer, because it is not the students who are moved forward more rapidly, but rather the opportunities they are provided that are accelerated. As such, they prefer the phrase "appropriate developmental placement" (Lubinski & Benbow, 2000, p. 138). Grade skipping and early admission into kindergarten or college are perhaps the most commonly known forms of academic acceleration, but recent reviews suggest there are as many as 20 forms of acceleration (Assouline et al., 2015), although many of these forms are variations of a similar practice. For example, early entrance into kindergarten, first grade, and middle school, high school, or college are considered three different forms of acceleration. Other forms of acceleration include self-paced instruction, subject-specific acceleration, curriculum compacting, dual enrollment, credit by examination, and early graduation.

Researchers have described five primary dimensions on which acceleration practices differ from typical educational experiences: pacing, salience, peers, access, and timing (Southern & Jones, 2015). The pacing dimension refers to the rate at which material is taught. Although it may seem counterintuitive, not all forms of acceleration offer more rapid pace than nonacceleration. For example, grade skipping, early entrance, and subject-specific acceleration do not change the pace of learning; they shift the age at which learning happens. The salience dimension has to do with the extent to which the intervention is observed by others, especially other students. Subject-specific acceleration, which could require students to physically leave one classroom for another, is more salient to the accelerated student and to peers.

The peers dimension refers to the extent to which students are separated from their same-age peers. Concerns about relative immaturity and separation from same-age peers are often raised by educators, parents, and students as a concern about acceleration (Southern, Jones, & Fiscus, 1989). Numerous studies have investigated the peer dimension of acceleration and generally reported not only no harm but also small to moderate social–emotional *benefits* of academic acceleration (Pressey, 1955; Rogers, 2015; Steenbergen-Hu & Moon, 2011). These findings do not indicate that there are never any social–emotional problems associated with acceleration; rather, they suggest that such problems are the exception, not the rule. The access dimension concerns the availability of acceleration options. Geographic and financial barriers may limit access to some forms of acceleration, though technological advancement may minimize the former while straining the latter. Finally, the timing dimension refers to the chronological age of students when the accelerative interventions are offered.

Estimating the prevalence of current acceleration opportunities is difficult given the varied state policies on gifted programs (McClain & Pfeiffer, 2012). A recent national survey of gifted programming at the elementary, middle, and high school levels (Callahan, Moon, & Oh, 2013) reported that districts that responded to the survey said that 90.7% offered Advanced Placement courses, 86.9% offered dual enrollment, and 13.1% offered International Baccalaureate. However, this survey only reports the percentage of responding districts that offer such programs, not the percentage of students who participate in the programs. Similarly, 68.2% of middle schools reported offering subject-specific acceleration, and 48.3% reported offering grade skipping. Callahan et al.'s (2013) survey of elementary schools asked how "most services" were delivered instead of the "check all that apply" option used in the middle and high school survey. With the change in framing of the question, only 1.7% of districts reported subject-specific acceleration and 0.2% (a single district) reported whole-grade acceleration as the primary form of service for gifted students.

History of Acceleration

In the early 20th century, the rationale for providing acceleration was rooted in the psychological findings of individual differences to help better match students and potential career planning through the creation of honors courses, individual work, interaction with mentors, and greater flexibility over how time is spent (Seashore, 1922). In Terman's (1925) classic *Genetic Study of Genius*, 84.5% of boys and 82.5% of girls were considered accelerated because they had started first grade prior to turning 6 years old. These students' teachers rated their work quite highly, particularly in areas that required abstract thought (e.g., debate, composition), but only slightly above average in areas where manual dexterity was measured (e.g., penmanship, painting). Terman noted that not all gifted students excelled; some performed poorly because they failed to complete daily assignments out of lack of interest, whereas others appeared to irritate their teachers.

In the 1940s, academic acceleration in the form of lengthening the school year increased in popularity in the United States as part of the nation's war effort. Then, following the war, the influx of older students (often war veterans) into colleges and universities led to continued demand for accelerated practices so that careers could begin sooner (Pressey, 1946). Skeptics have noted that the use of acceleration in the form of year-round training of medical, dental, and engineering military members may have been effective as part of an emergency war effort, but that under normal circumstances, acceleration was often rife with dropouts and the opportunity cost of missing out on essential course content as well as important life experiences (Smith, 1945). Perhaps in response to skeptics, Pressey (1955, p. 127) concluded that "numerous studies are practically unanimous in showing that able children can enter earlier and progress more rapidly than the average child, without harm and often with gain in regard to realized abilities and social adjustment"; he considered grade skipping to be among the "worst" methods of acceleration, preferring early entrance to first grade, compacting multiple years of elementary school into less time, and giving credit via examination for college courses.

Others have advocated for radical acceleration (accelerating the equivalent of 3 or more years) for some of the most able students (Gross, 2004). One example of extreme acceleration in what Stanley (2000) admitted to be an educational "stunt" was when 75 students in the Study of Mathematically Precocious Youth who had been identified as especially strong in math were taught a year's worth of algebra in a single day. This intervention was resource intensive with an extremely select sample of academically talented students, but serves as existence proof for what the extreme outer envelope of acceleration can look like. Additionally, historiometric analyses of the development of eminent individuals (e.g., Cox, 1926) reveal that many experienced some form of radical acceleration in their field when they were young (Gross, 1992). Colangelo, Assouline, and Gross (2004) suggested that the failure to implement acceleration interventions is due to *a nation deceived* into believing acceleration is ineffective or harmful, with a more recent report reframing the discussion as *a nation empowered* to accelerate students (e.g., Assouline et al., 2015).

The Present Study

Meta-analysis, or first-order meta-analysis, is a quantitative research review method for combining and comparing the results from multiple primary studies to generate a synthesis of the outcomes on a given topic or relationship (Glass, McGaw, & Smith, 1981). A second-order meta-analysis is a meta-analysis of a number of methodologically comparable existing first-order meta-analyses that examined similar issues or relationships on a given topic (Cooper & Koenka, 2012; Schmidt & Oh, 2013). Second-order meta-analyses are also known as overviews of reviews, systematic reviews of reviews, umbrella reviews, meta-meta-analyses, and meta-analyses of meta-analyses (Polanin, Maynard, & Dell, 2016). Second-order meta-analyses have increased in importance, as meta-analysis has become a widely accepted research method in education over the past two decades, with the relationship between a second-order meta-analysis and related meta-analyses being quite similar to that of a meta-analysis to primary studies. This relatively new form of scholarship has been most widely used in the medical and health sciences.

Dozens of meta-analyses on the impact of ability grouping and acceleration on students' academic achievement have been conducted from the 1980s to present. These meta-analyses have drawn great attention from education researchers and practitioners. For example, Slavin's (1987, 1990, 1993) syntheses of studies on ability grouping have been cited 1,188, 965, and 151 times, respectively, according to Google Scholar as of August 2016. However, no second-order meta-analysis has been conducted to integrate and synthesize these existing meta-analyses. Second-order meta-analyses can serve some important purposes, such as (a) summarizing evidence from more than one meta-analysis, (b) comparing findings and resolving discrepancies among existing meta-analyses, (c) reexamining the credibility and validity of the conclusions of existing meta-analyses with a fresh perspective and new advancements in meta-analysis techniques, and (d) identifying research gaps and future inquiry directions.

Research Questions

In line with the purposes mentioned above, the two second-order meta-analyses in this article address five specific research questions.

> **Research Question 1:** What are the effects of ability grouping and acceleration on K–12 students' academic achievement as shown by integrating findings of existing meta-analyses?
> **Research Question 2:** Does ability grouping have differential impacts on students of different ability levels (e.g., high, medium, and low ability)?
> **Research Question 3:** What are the discrepancies and commonalities in the methods and findings across different meta-analyses?
> **Research Question 4:** Do meta-analyses of different methodological quality show differential effects?

Research Question 5: What are the effects of ability grouping when only the highest quality of research evidence is considered?

Method

Meta-Analysis Inclusion/Exclusion Criteria

As with a typical first-order meta-analysis, the current second-order meta-analyses established a set of inclusion or exclusion criteria as below to determine eligible first-order meta-analyses to include. These included the following:

1. It employed methods of meta-analysis or quantitative synthesis to aggregate research findings, often through the calculation of effect sizes (ESs; Cooper & Hedges, 1994). The term, meta-analysis, was used to represent several similar expressions such as quantitative synthesis, best-evidence synthesis, and meta-analytic reviews.
2. It focused on the academic impact of ability grouping, acceleration, or both.
3. It included studies that had both treatment and control groups so that standardized mean differences were calculable. Reviews of research literature that did not report ESs were excluded.
4. It reported academic achievement outcomes of ability grouping or acceleration interventions. If a meta-analysis included both cognitive and affective outcomes, only the cognitive outcomes were used in the current second-order meta-analyses. Meta-analyses that focused only on nonacademic outcomes (such as social–emotional outcomes) were excluded (e.g., Kent, 1992).
5. It had a written document available, either published or unpublished.
6. It was available in English.

Conducting the Search and Determining Eligibility

Three procedures were used to search for eligible meta-analyses: (a) searches of electronic databases, including ERIC, PsycINFO, Academic Search Premier, EconLit with Full Text, PsycARTICLES, SocINDEX with Full Text, Education Full Text, and Academic Search Complete; (b) web searches using Google and Google Scholar engines; and (c) manual examinations of references and bibliography lists of the relevant literature. The searches combined two sets of search terms. One set was used to identify meta-analyses of substantive relevance, that is, meta-analyses that focused on either ability grouping or acceleration and met the inclusion criteria. Specifically, the set of search terms used to identify ability grouping meta-analyses included grouping, ability grouping, tracking, streaming, setting, sorting, classroom organization or composition, and classroom assignment. To search for acceleration meta-analyses, we conducted our search with terms associated with 20 types of acceleration practices identified in *A Nation Empowered* (Assouline et al., 2015), such as grade skipping, early college entrance, curriculum compacting, and advanced placement.

The other set of search terms used to locate syntheses on ability grouping and acceleration included meta-analysis, meta-analytic, quantitative synthesis, best-evidence synthesis, and systematic reviews. The searches were conducted with the use of truncation (such as “”), wildcard (e.g., *), Boolean operators (AND, OR, and NOT), and limiting commands to ensure search sensitivity and precision. No time restriction was applied to the search, which was completed in January 2016.

We screened thousands of article titles and abstracts in the search process. We determined the eligibility of meta-analyses based on whether the actual instructional practices in the relevant literature met our definition of ability grouping and acceleration practices. Although numerous narrative reviews of ability grouping and acceleration exist, the number of meta-analyses of these issues is limited. We initially identified 30 meta-analyses of ability grouping and 10 of acceleration. Two authors read the full text of the 40 meta-analyses and determined that 13 meta-analyses on ability grouping and six on acceleration met the inclusion criteria of the current second-order meta-analyses.

Coding

Each meta-analysis was coded with a detailed coding form modified from the coding sheet designed by Ahn, Ames, and Myers (2012). The coding form covered the major characteristics of eligible meta-analyses, such as (a) general features of the meta-analyses, including publication status, definitions of ability grouping or acceleration, types of ability grouping or acceleration, main instructional features of intervention and comparison groups, and key research questions; (b) methodological features, including inclusion and exclusion criteria, main search terms, total number of studies reviewed, major research design of the included studies, outcome measures, moderator analyses as applicable, and interpretations of results; and (c) ES, ES index, ES calculation and extraction methods, number of ESs reported, integration of ESs (e.g., mean or median), and main conclusion of the meta-analyses.

To pilot the coding form, the first and second authors independently double-coded several meta-analyses. They then met to check the coding agreement and resolved discrepancies. Examples of causes of pilot coding disagreement included discrepancies in the interpretation of the conceptual and operational aspects of the studies or overlooking information that was hard to find in the meta-analyses. Through the pilot coding, the first and second authors fine-tuned the coding form and reached consensus on how to handle similar issues that might arise in later coding. They then proceeded to code the remaining ability grouping and acceleration meta-analyses, respectively. Both coders have more than 10 years of experiences in education research. Most important, each coded their respective meta-analyses at least two times to ensure the coding accuracy and reliability. Additionally, the first author examined the coding of acceleration meta-analyses prior to conducting the data analysis.

Assessing Study Overlap and Methodological Quality of Meta-Analyses

Polanin et al. (2016) noted that overview authors must be aware of study overlap across included reviews. Cooper and Koenka (2012) summarized various strategies to handle overlap in overviews, such as selecting the review that is most rigorous, or is the most recent and published in a reputable journal. However, it is not clear which approach is the most appropriate and each approach may be justifiable depending on the overview (Cooper & Koenka, 2012). In this study, we focused on assessing the amount of overlap across meta-analyses. To examine study overlap, we formed a reference group that consisted of 68 nonduplicated primary studies from three meta-analyses by Slavin (1987, 1990, 1993) because they appeared to be the most rigorous, contained the most complete description of reviewed studies, and had been widely cited meta-analyses of this issue in the past two decades. We then compared the references of each of the remaining 10 meta-analyses with the 68 reference studies to calculate the number of overlapping studies. We also cross-referenced among the 13 ability grouping meta-analyses to assess additional overlapping information. For acceleration meta-analyses, we used the reference list of Rogers (1991) as a reference group because it was the most extensive. We then compared the references of the remaining five acceleration meta-analyses to compute the number of overlapping studies. Finally, the percentage of overlapping studies was calculated by dividing the number of overlapping studies with the total number of identifiable studies included in a meta-analysis.

To assess the methodological quality of meta-analyses, we utilized an empirically developed and validated instrument, AMSTAR (Shea et al., 2009). AMSTAR has good agreement (*M* kappa = 0.70), reliability, construct validity, and feasibility to assess the methodological quality of a wide variety of systematic reviews. AMSTAR consists of 11 items. Each item can be rated as *yes*, *no*, *can't answer*, or *not applicable*. Each *yes* response was assigned 1 point. The sum of earned points from the 11 items represented the methodological quality score for each meta-analysis. We categorized meta-analyses with 5 points or fewer as low in methodological quality, those with 6 to 9 points as moderate, and those with 10 to 11 points as high in methodological quality.

Second-Order Meta-Analyses

We integrated mean ESs across 13 meta-analyses of ability grouping and 6 of acceleration. Additionally, we carried out a mini first-order meta-analysis of randomized experimental studies of ability grouping. Researchers in health sciences, education, and psychology have commonly used these approaches for second-order meta-analyses (Cooper & Koenka, 2012). We elaborate on each approach below.

Integrating Mean Effect Sizes Across Meta-Analyses

Integrating mean ESs across meta-analyses consisted of three steps. The first step was extracting mean ESs from the original first-order meta-analyses. The original mean ESs were in line with the indices that the original meta-analyses employed. The majority of the 13 ability grouping meta-analyses used Cohen's *d* or an equivalent of Cohen's *d* (often called "standard deviation") as ES indices, except Goldring (1990) who used ESs equivalent to Hedges' *g* and Noland (1986) who used Glass' Δ. All acceleration meta-analyses used Cohen's *d* except Steenbergen-Hu and Moon (2011) who reported Hedges' *g*. Regardless, all were standardized mean differences and they were generally computed as the difference between the experimental and control means divided by the control group's standard deviation in the meta-analyses.

In cases where the original meta-analyses reported median or nonaggregated ESs rather than means, we computed relevant mean ESs from available ESs of individual primary studies reported in the original meta-analyses. For example, Slavin (1987) reported an overall median ES of zero for between-class grouping. To obtain a corresponding original mean ES, we averaged the 15 ESs that were associated with the 14 reviewed primary studies as reported in Table 1 of Slavin (1987). This calculation yielded an original mean ES of −0.56 (see Table 2). The Supplementary Appendix (available in the online version of the journal) includes all relevant notes describing how particular original mean ESs were computed or extracted from the original meta-analyses.

The second step of integration involved converting the original mean ESs into Hedges' *g*—a common index—so that integrating the outcomes of different meta-analyses became possible. Hedges' *g* can reduce the bias that may arise when the sample size is small (i.e., $n < 40$; Glass et al., 1981; Hedges, 1981), in comparison with some other standardized mean differences (Borenstein, Hedges, Higgins, & Rothstein, 2009). A positive Hedges' *g* indicates that ability grouped students or accelerated students outperformed their comparison peers.

The conversion was conducted using the Comprehensive Meta-Analysis (CMA) software (Borenstein, Hedges, Higgins, & Rothstein, 2006), which has a built-in function that operates on the basis of original mean ESs and associated standard errors (*SE*s). However, *SE*s were available for only 4 of the 66 original mean ESs for ability grouping meta-analyses (see Table 2, Footnote b, for detailed information). We therefore used the average of these four available *SE*s ($SE = 0.09$) in converting each original mean ES to a Hedges' *g*. Similarly, we employed the average of 11 *SE*s available in acceleration meta-analyses ($SE = 0.23$) for such conversions. We acknowledge that this is a compromised imputation, so we also tested whether Hedges' *g*s yielded through such conversions varied substantially when they were calculated based on available alternative information such as confidence intervals (CIs; see Table 8). Results showed that converted Hedges' *g*s were relatively conservative relative to the original mean ESs.

The third step was integrating the mean ESs across meta-analyses. The integration was conducted with the CMA software using mixed-effects models and the Hedges' *g*s of individual meta-analyses as the unit of analysis. An integrated Hedges' *g* is the average of the Hedges' *g*s across associated original meta-analyses. Each of the contributing Hedges' *g* was weighted using the Hedges and Vevea (1998) method of weighting by inverse variances, so that the integrated Hedges' *g* was not overly influenced by any single ES. Moderator analyses that were analogous to the analysis of variance were conducted to examine whether the integrated mean ESs differed by the methodological quality level of meta-analyses.

Assess Publication Bias

As applicable, analyses were also conducted to assess whether publication biases existed and their influences on the integrated mean ESs. Publication bias exists when studies with nonfavorable results, particularly those with negative or nonsignificant findings, are less likely to be published or accessible to researchers than those with positive or significant outcomes (Rothstein, Sutton, & Borenstein, 2005). Existence of publication bias may lead to biased estimates (usually overestimates)

(Text continues on p. 868.)

TABLE 1
Major features of ability grouping meta-analyses

Meta-analysis[a]	No. of studies[b]	Study coverage		Subjects	Grades	Original grouping type[c]	Recategorized grouping type[d]	Comparison condition[e]
		Earliest	Newest					
Kulik and Kulik (1982)	51	9 studies before 1951	6 studies between 1971 and 1980	ng	Secondary	Between-classes ability grouping	Between-class grouping	Ungrouped classes
Kulik and Kulik (1984a)	28	ng	ng	ng	Elementary	Between-classes ability grouping	Between-class grouping	Ungrouped classes
Kulik (1985)	78	ng	ng	ng	Elementary, secondary	Homogeneous grouping	Between-class grouping	Heterogeneous grouping
Noland (1986)	≤50	1967	1983	ng	K–12	Ability grouping	Between-class grouping	Heterogeneous grouping
Kulik and Kulik (1987)	109	1928	1984	ng	Grades 1–12	Between-class comprehensive grouping	Between-class grouping	Heterogeneous grouping
						Between-class special grouping of talented students	Special grouping for the gifted	Heterogeneous grouping
						Joplin Plan programs	Cross-grade subject grouping	Heterogeneous grouping
						Within-class grouping students of all abilities	Within-class grouping	Heterogeneous grouping
						Within-class enriched programs	Special grouping for the gifted	Heterogeneous grouping
Slavin (1987)	43	1932	1985	24,534, plus 4 schools	Elementary	Comprehensive ability-grouped classes	Between-class grouping	Heterogeneously grouped classes
						Ability grouping for selected subjects (i.e., reading and mathematics)	Cross-grade subject grouping	Heterogeneously grouped classes

(continued)

TABLE 1 (continued)

Meta-analysis[a]	No. of studies[b]	Study coverage		Subjects	Grades	Original grouping type[c]	Recategorized grouping type[d]	Comparison condition[e]
		Earliest	Newest					
						Joplin Plan (and Joplin-like nongraded plans)	Cross-grade subject grouping	Heterogeneously grouped classes
						Within-class ability grouping	Within-class grouping	Heterogeneously grouped classes
Henderson (1989)	4	1985	1987	20,681	Elementary	Homogeneous ability grouping	Between-class grouping	Heterogeneous grouping
Goldring (1990)	18	1959	1985	ng	Grades 3–12	Gifted students in special, homogeneous classes	Special grouping for the gifted	Gifted students in regular heterogeneous classes
Slavin (1990)	29	1927	1986	25,289, plus 19 schools	Secondary	Ability-grouped classes	Between-class grouping	Heterogeneously grouped classes
						Within-class ability grouped classes	Within-class grouping	Heterogeneous assignment within-class
Kulik and Kulik (1992)[f]	124	ng	ng	ng	Grades 1–12	Multilevel classes	Between-class grouping	Mixed-ability classes
						Cross-grade grouping (e.g., the Joplin Plan)	Cross-grade subject grouping	Mixed-ability classes
						Within-class grouping	Within-class grouping	Mixed-ability classes
						Enriched classes for the gifted and talented	Special grouping for the gifted	Mixed-ability classes
Slavin (1993)	27	1927	1991	25,718, plus 12 schools	Middle Grades (6–9)	Ability-grouped classes	Between-class grouping	Heterogeneously grouped classes

(continued)

TABLE 1 (continued)

Meta-analysis[a]	No. of studies[b]	Study coverage		Subjects	Grades	Original grouping type[c]	Recategorized grouping type[d]	Comparison condition[e]
		Earliest	Newest					
						Within-class ability grouped classes	Within-class grouping	Heterogeneous assignment within-class
Mosteller et al. (1996)	15	1960	1985	3,821	Grades 1–11	XYZ groupings	Between-class grouping	Whole-class grouping
						The Joplin Plan	Cross-grade subject grouping	Whole-class grouping
						Within-class grouping	Within-class grouping	Whole-class grouping
Lou et al. (1996)	66	1922	1994	16,073	Elementary, secondary, postsecondary	General within-class grouping	Within-class ability grouping	No grouping
						Within-class homogeneous grouping	Within-class grouping	Within-class heterogeneous grouping

Note. ng = not given.

[a] The meta-analyses are listed in a chronological order.

[b] The number of included studies only refers to studies of academic achievement. Several meta-analyses include both studies of academic achievement and noncognitive outcomes.

[c] The “original grouping type” in this column are the original grouping categories or terminologies used in the original first-order meta-analyses.

[d] The “recategorized grouping type” in this column denotes the related four main grouping types as categorized in this second-order meta-analysis after recoding the original grouping categories used in the original first-order meta-analyses (i.e., those listed the column for “original grouping category”).

[e] The “comparison condition” in this column is the comparison groups or conditions described in the original first-order meta-analyses.

[f] Kulik and Kulik (1992) meta-analysis focused on ability grouping but it also reviewed a small group of studies of the impact of acceleration on gifted students. We chose to omit these findings from this second-order meta-analysis on ability grouping.

TABLE 2

Effects of between-class ability grouping on academic achievement

Outcomes of the individual original meta-analyses								
Meta-analysis	Student group	No. of studies (ESs)	Original mean ES^{a}	Hedges' g^{b}	Standard error^{c}	95% CI	p	Narratives/interpretations in the original meta-analysis
Kulik and Kulik (1982)	Overall	51	0.10[1]	0.10	0.09	[−0.08, 0.27]	.27	"The average ES of .10 implies that in a typical class, performance of ability-grouped students was raised approximately one-tenth of a standard deviation unit." (p. 420)
	Medium ability (i.e., representative of population)^{d}	33	0.02	0.02	0.09	[−0.15, 0.19]	.82	
	Low ability (i.e., academically deficient)	4	0.02	0.02	0.08	[−0.14, 0.17]	.82	
Kulik and Kulik (1984a)	Overall	28	0.19[2]	0.19	0.09	[0.01, 0.36]	.03	"This means that in the typical study, grouping raised student grade-equivalent scores by approximately 2 months. It is also equivalent to raising student achievement scores from the 50th to the 58th percentile." (p. 5)
	Medium ability (i.e., more representative populations)	19	0.07	0.02	0.09	[−0.10, 0.24]	.44	"In studies of programs for more representative populations, grouping raised achievement tests scores by .07 standard deviations." (p. 5)
Kulik (1985)	Overall	78	0.15[3]	0.15	0.09	[−0.03, 0.32]	.10	"That is, the average effect of homogeneous grouping in the 78 studies was to raise examination scores by 0.15 standard deviations. This is a small effect, equivalent to an increase of about 1.5 months on a grade equivalent scale, or an increase from the 50th to the 56th percentile on a percentile scale." (p. 4)
	Medium ability (i.e., students representing a full range of ability)	≤74	0.12	0.12	0.09	[−0.06, 0.29]	.18	". . . in general, effects of XYZ grouping are less impressive than are the results of honors program." (p. 5)
	Low ability (i.e., remedial programs)	4	0.14	0.12	0.08	[−0.03, 0.28]	.12	"The average effect in the 4 studies of remedial grouping was to raise achievement scores of slow learners by 0.14 standard deviations." (p. 5)

(continued)

TABLE 2 (continued)

Outcomes of the individual original meta-analyses								
Meta-analysis	Student group	No. of studies (ESs)	Original mean ES[a]	Hedges' *g*[b]	Standard error[c]	95% CI	*p*	Narratives/interpretations in the original meta-analysis
Noland (1986)	Overall	≤50 (272 ESs)	0.01	0.01	0.09	[−0.17, 0.18]	.91	"Thus, the hypothesis that the cognitive outcome scores of students when they were ability grouped for instruction and the cognitive outcome scores of students who were not ability grouped do not differ cannot be rejected." (p. 133)
	High ability	≤50 (82 ESs)	0.16	0.16	0.09	[−0.02, 0.33]	.08	". . . When only cognitive measures are considered, the effects of ability grouping are positive for both high and low ability students (+ 0.16, $n = 82$, and +0.18, $n = 72$, respectively) but quite negative, −0.45, $n = 36$), for average ability students." (p. 148)
	Medium ability	≤50 (36 ESs)	−0.45	−0.45	0.09	[−0.62, −0.27]	<.01	
	Low ability	≤50 (72 ESs)	0.18	0.18	0.09	[0.00, 0.35]	.05	
Kulik and Kulik (1987)	Overall	49	0.06	0.06	0.09	[−0.12, 0.23]	.50	"The average effect in these 49 studies was to raise student performance by 0.06 standard deviation. This effect is significantly greater than zero, $t(48) = 2.04$, $p < .05$, but it is clearly very small in size." (p. 24)
	High ability	40 (40ESs)	0.12	0.12	0.09	[−0.06, 0.29]	.18	"The average effect size was 0.12 for students in the high-ability classes." (p. 24)
	Medium ability	33 (33ESs)	0.04	0.04	0.09	[−0.13, 0.21]	.66	"The average effect size was 0.04 for students in the middle-level classes." (p. 24)
	Low ability	39 (39 ESs)	0	0.00	0.09	[−0.17, 0.17]	1.00	"The average effect size was 0.00 for students in the low-ability classes." (p. 24)
Slavin (1987)	Overall	14 (15 ESs)	−0.56[4]	−0.54	0.09	[−0.71, −0.37]	<.01	"Inspection of Table 1 clearly indicates that the effects of comprehensive ability grouped class assignment on student achievement are zero." (p. 301)

(continued)

TABLE 2 (continued)

Outcomes of the individual original meta-analyses								
Meta-analysis	Student group	No. of studies (ESs)	Original mean ES[a]	Hedges' g[b]	Standard error[c]	95% CI	p	Narratives/interpretations in the original meta-analysis
Henderson (1989)	Overall	4	−0.34	−0.30	0.08	[−0.45, −0.14]	<.01	"The average effect of heterogeneous grouping in the four studies was to raise student scores on achievement test by .34 standard deviation. This means that heterogeneous grouping raised student grade-equivalent scores by approximately three months." (pp. 48–49)
	High ability	2	0.039	0.02	0.05	[−0.08, 0.12]	.66	"An average effect size of +0.039 was found in favor of homogeneous grouping. The effect of homogeneous grouping was to raise students' scores on the achievement test by 0.039 standard deviations." (p. 52)
	Medium ability (i.e., combined average and low group)	2	−0.00025[5]	−0.00014	0.05	[−0.10, 0.10]	1.00	"The effect of heterogeneous grouping on average and low ability students in this study was to raise student scores on achievement tests by .00025 standard deviations." (p. 54)
Slavin (1990)	Overall	29	−0.03[6]	−0.03	0.09	[−0.20, 0.14]	.74	"The median effect size for the 20 studies from which effect sizes could be estimated was −.02, and none of the 9 additional studies found statistically significant effects. Counting the studies with nonsignificant differences as though they had effect sizes of .00, the median effect sizes for all 29 studies would be .00." (p. 484)
	High ability	15	−.02[7]	−0.02	0.09	[−0.19, 0.15]	.82	"Across the 15 studies from which effect sizes could be computed, the median effect size was +.01 for high achievers, −.08 for average achievers, and −.02 for low achievers. Effects of this size are indistinguishable from zero, and if all the nonsignificant differences found in studies from which effect sizes could not be computed are counted as effect sizes of .00, the median effect size for each level of student becomes .00." (p. 485)
	Medium ability	15	−0.07[8]	−0.07	0.09	[−0.24, 0.10]	.44	
	Low ability	15	−0.03[9]	−0.03	0.09	[−0.20, 0.14]	0.74	

(continued)

TABLE 2 (continued)

		Outcomes of the individual original meta-analyses						
Meta-analysis	Student group	No. of studies (ESs)	Original mean ES[a]	Hedges' *g*[b]	Standard error[c]	95% CI	*p*	Narratives/interpretations in the original meta-analysis
Kulik and Kulik (1992)	Overall	51	0.03	0.03	0.09	[−0.15, 0.20]	.74	"The average effect size in all programs was 0.03. The effect is not large enough to be considered statistically different from zero." (p. 74)
	High ability	36	0.10	0.10	0.09	[−0.08, 0.27]	.27	"The average effect size was 0.10 for higher aptitude, −0.02 for middle aptitude, and −0.01 for lower aptitude students." (pp. 74–75)
	Medium ability	36	−0.02	−0.02	0.09	[−0.19, 0.15]	.82	
	Low ability	36	−0.01	−0.01	0.09	[−0.18, 0.16]	.91	
Slavin (1993)	Overall	27	−.01[10]	−0.01	0.09	[−0.18, 0.16]	.91	". . . the median effect size for all 27 studies would be .00." (p. 539)
	High ability	14	0.01[11]	0.01	0.09	[−0.16, 0.18]	.91	"Across the 14 studies from which effect sizes could be computed, the median effect size was +.09 for high achievers, −.07 for average achievers, and −.05 for low achievers." (p. 545)
	Medium ability	14	−0.07[12]	−0.07	0.09	[−0.24, 0.10]	.44	
	Low ability	14	−0.02[13]	−0.02	0.09	[−0.19, 0.15]	.82	
Mosteller et al. (1996)	Overall	10	0	0.00	0.09	[−0.17, 0.17]	1.00	"Overall, results of the ten studies suggest that XYZ grouping, on average, does not have much effect on achievement." (p. 807)
	High ability	10	0.08	0.08	0.09	[−0.09, 0.25]	.37	"Results from the ten studies suggest that XYZ grouping seems modestly preferable to whole-class grouping for high-skill students. In contrast, medium-and low-skill students may learn a little more with whole-class instruction that with skill grouping." (p. 806)
	Medium ability	10	−0.04	−0.04	0.09	[−0.21, 0.13]	.66	
	Low ability	10	−0.06	−0.06	0.09	[−0.23, 0.11]	.50	

(continued)

TABLE 2 (continued)

	Integrated outcomes across meta-analyses								
	Meta-analytic statistics					Heterogeneity across meta-analyses			
Grouping type/student group	No. of meta-analyses	Integrated Hedges' g^e	Standard error	95% CI	p	$Q_b{}^f$	df	p	I^{2g}
Between-class grouping									
Overall	11	−0.03	0.07	[−0.16, 0.09]	.62	59.87	10	<.01	83.30
High ability	7	0.06	0.03	[−0.001, 0.11]	.06	3.56	6	.74	0.00
Medium ability	10	−0.04	0.05	[−0.13, 0.05]	.40	27.44	9	<.05	67.20
Low ability	8	0.03	0.03	[−0.03, 0.09]	.37	6.28	7	.51	0.00

Note. CI = confidence interval; ES = effect size; *df* = degrees of freedom; *SE* = standard error. Notations 1 to 13 describe how particular original mean ESs were computed or extracted from the original meta-analyses. They are included as notes in the Supplementary Appendix (available in the online version of the journal). There were cases in several meta-analyses where a small number of studies were reviewed as a category but the outcomes were not aggregated quantitatively and only described narratively. For example, although Slavin (1987, 1990) meta-analyses focused primarily on between-class grouping, they also reviewed studies of within-class grouping and described the outcomes narratively. No aggregated mean ESs were extracted in such cases.

[a] The *original mean* ES denotes the mean ESs (i.e., standardized mean differences) extracted from the original first-order meta-analyses. These original mean ESs are in line with whichever ES indices the original meta-analyses employed, of which the majority of them were Cohen's *d* or equivalent of Cohen's *d*.

[b] The Hedges' *g* denotes a mean ES converted or recalculated from the original mean ES through the Comprehensive Meta-Analysis software with a built-in function. With this function, a Hedges' *g* was computed after entering an original mean ES, a standard error associated with the original mean ES, and the respective sample size of the intervention and comparison group. It is important to note that associated standard errors were available for only 4 of the 66 original mean ESs. Three of the standard errors were associated with the original mean ESs for between-class grouping in Kulik and Kulik (1982), corresponding to overall ($SE = 0.045$), medium ability ($SE = 0.04$), and low-ability students ($SE = 0.18$). One standard error was associated with the original mean ES for special grouping for the gifted in Kulik and Kulik (1992), corresponding to the gifted students overall ($SE = 0.11$). Therefore, we used the average ($SE = 0.09$) of these four available *SE*s as a common *SE* in converting the original mean ESs to the Hedges' *g*s.

[c] The *standard error* is associated with each of the Hedges' *g* after a conversion from the corresponding original mean ES.

[d] The contents enclosed in the parentheses are terms/phrases of the original meta-analyses.

[e] An *integrated Hedges' g* is the combined average of the *Hedges' g*s across associated original meta-analyses.

[f] Q_b denotes the value of total between-meta-analysis variance.

[g] I^2 and Q_b are algebraically related. It is another indicator of heterogeneity, ranging from 0 to 100 (%). According to Higgins and Green's (2011) rough guide for I^2 interpretation, an I^2 of 0% to 40% indicates "likely not important," 30% to 60% indicates "possible moderate heterogeneity," 50% to 90% indicates "possible substantial heterogeneity," and 75% to 100% indicates "considerable heterogeneity."

of an effect. The current second-order meta-analyses assessed publication bias through producing a funnel plot (Peters, Sutton, Jones, Abrams, & Rushton, 2008) and conducting trim-and-fill analysis (Duvall & Tweedie, 2000). We chose this relatively conventional approach over some recently developed meta-regression methods such as the PET model (Stanley, 2005), the precision-effect estimate with standard error (PEESE) model (Stanley & Doucouliagos, 2007), and PET–PEESE procedure (Stanley & Doucouliagos, 2014) because our second-order meta-analyses only involved a small number of meta-analyses, whereas meta-regression approaches often require more than a few estimates (Stanley & Doucouliagos, 2014).

A Mini Meta-Analysis of 12 Randomized Studies

We conducted a mini meta-analysis of 12 randomized controlled trials (RCTs) that had been reviewed meta-analytically previously to examine how the effects of ability grouping manifest when only the highest quality of research evidence is considered (Research Question 5). It was our intention that this mini meta-analysis would serve as a unique opportunity to reexamine the effects of ability grouping reported in previous meta-analyses at a more granular level. We focused on RCTs because they are viewed as the "gold standard" of education research according to the What Works Clearinghouse (Whitehurst, 2003). We identified a total of 15 RCTs that had been reviewed in the three meta-analyses by Slavin (1987, 1990, 1993), including the Bicak (1964) study, which was an updated version of the Bicak (1962) dissertation. Intending to limit the scope of this mini meta-analysis, we used these 15 RCTs as a sample of randomized studies on ability grouping. We were able to retrieve the full texts of 13 RCTs in the end.

Of the 13 studies retrieved, the study by Marascuilo and McSweeney (1972) was excluded because ESs were not calculable. The article by Hillson, Jones, Moore, and Van Devender (1964) was excluded because it shared the same set of data as Jones, Moore, and Van Devender (1967). The article by Slavin and Karweit (1985) actually consisted of two experimental studies (Experiments 1 and 2). We therefore treated each experiment as a separate study. After exclusions, the mini meta-analysis was conducted on a total of 12 studies with Hedges' *g* as an ES index. The analyses were conducted with mixed-effects models, focusing on computing mean ESs across relevant primary studies. No moderator analyses nor assessments of publication bias were conducted for the mini meta-analysis.

Results on Ability Grouping

Descriptive Information

Table 1 presents the key features of the 13 ability grouping meta-analyses that met all inclusion criteria. There were 172 unique primary studies in the 13 meta-analyses after eliminating the duplicated studies from the total identifiable primary studies. The earliest primary study was published in 1922 and the most recent was published in 1994. Of the 172 unique primary studies, 5 appeared in the 1920s, 6 in the 1930s, 3 in the 1940s, 12 in the 1950s, 65 in the 1960s, 23 in the 1970s, 38 in the 1980s, and 20 studies appeared in the 1990s. The majority (85%) appeared between the 1960s and 1990s, and almost 38% of the studies were conducted in the 1960s. The number of participants ranged from 3,821 to

25,718. The duration of ability grouping ranged from 1 week to 7 years. Ability grouping was most frequently used for learning math, reading, science, language arts, and social studies.

Study Overlap Between Ability Grouping Meta-Analyses

Supplementary Table S1 (available in the online version of the journal) presents the number of primary studies in each meta-analysis, the number of studies overlapping the reference group, and the percentage of overlap. Several findings are worth noting. First, it is quite likely that most of the primary studies covered in two latter meta-analyses by the Kuliks (i.e., 1987, 1992) overlapped with those in their three early meta-analyses (i.e., Kulik, 1985; Kulik & Kulik, 1982, 1984a), although it was not possible to calculate the exact amount of overlap, as the references of the majority of the primary studies included in the five meta-analyses by the Kuliks were not provided. Second, studies included in the Kulik meta-analyses largely overlapped with those included in the meta-analyses by Slavin (1987, 1990), as Kulik and Kulik (1992) stated that most of the 124 studies included their meta-analysis came from their earlier meta-analyses and from Slavin's studies. Third, 13 of the 15 (86.7%) experimental studies reviewed in the Mosteller et al. (1996) meta-analysis overlapped with those reviewed in the Slavin meta-analyses. Fourth, only 5 (7.5%) of the 67 studies included in Lou et al. (1996) meta-analysis of within-class grouping overlapped with those in the Slavin's three meta-analyses. Last, none of the 18 studies in Goldring (1990) and 4 four studies in Henderson (1989) overlapped with those in the Slavin meta-analyses. In brief, the amount of study overlap in the 10 comparison meta-analyses and the reference meta-analyses ranged from 0% to 86.7%. Taken together, 10 of the 13 (76.9%) meta-analyses had at least moderate or high degree of study overlap.

Methodological Quality of Ability Grouping Meta-Analyses

Seven meta-analyses were rated as having low methodological quality (Henderson, 1989; Noland, 1986; and all five meta-analyses by the Kuliks), and six had moderate methodological quality (Goldring, 1990; Lou et al., 1996; Mosteller et al., 1996; Slavin, 1987, 1990, 1993; see Supplementary Table S2, available in the online version of the journal). No meta-analysis had high quality. The seven meta-analyses with low methodological quality had major weaknesses. For example, all five meta-analyses by the Kuliks lacked most of the information required for a typical a meta-analysis, such as specifications of study inclusion and exclusion criteria, literature search strategy, coding strategy, ES calculation and extraction, research design, and other substantial features of the included studies. Of the 13 meta-analysis, only Lou et al. (1996) assessed the likelihood of publication bias.

Four Major Types of Ability Grouping and Comparison Conditions

Meta-analyses used various terms for ability grouping. For instance, several meta-analyses used different terms to describe some types of grouping that are essentially between-class grouping, such as "comprehensive ability-grouped classes," "XYZ groupings," and "multilevel classes." Three meta-analyses labeled their topic of review as "homogeneous (ability) grouping" or "ability grouping."

We recoded the original grouping categories into four main types as described in the introduction: between-class grouping, within-class grouping, cross-grade subject grouping, and special grouping for the gifted. Eleven of the 13 (84.6%) ability grouping meta-analyses reviewed the academic effects of between-class grouping, five (38.5%) reviewed within-class grouping, four (30.8%) reviewed cross-grade subject grouping, and six (46.2%) reviewed special grouping for the gifted.

Twelve (92.3%) meta-analyses—that is, all except Goldring (1990)—not only reported the overall effects of ability grouping on students' achievement but also broke down the outcomes by high-, medium-, and low-ability students. For example, Slavin (1990) reported overall findings of the reviewed 29 studies of between-class grouping; it also reported the differential effects of between-class grouping by high, average, and low achievers. Despite variation in the terms used across meta-analyses, the core of all comparison conditions appeared to be quite similar: ability grouping was consistently compared with heterogeneous grouping in all included meta-analyses. We therefore categorized all these comparison conditions as "heterogeneous grouping."

Integrating Mean Effect Sizes Across 13 Meta-Analyses

Tables 2 to 5 present the main outcomes by ability grouping type. Each of these tables consists of two sections, labeled "Outcomes of Individual Original Meta-Analyses" and "Integrated Outcomes Across Meta-Analyses."

Outcomes of Individual Original Meta-Analyses

This section of Tables 2 to 5 contains information that was extracted from the individual first-order meta-analyses, corresponding to student groups, such as overall, high-, medium-, and low-ability students. Such information includes the original mean ES extracted from the individual meta-analyses, the number of primary studies or ESs that contributed to each original mean ES, and relevant narratives or interpretations quoted from the meta-analyses. A total of 66 original mean ESs were extracted from the 13 meta-analyses.

Also included in this section are Hedges' *g*s, which were converted from corresponding original mean ESs, *SE*s, 95% CIs, and *p* values. Simply averaging the 66 original mean ESs led to a mean of 0.22, whereas averaging the 66 Hedges' *g*s yielded a mean of 0.20. The values of 32 (48%) of the Hedges' *g*s were smaller than and those of 34 (52%) original mean ESs, making the Hedges' *g*s a slightly more conservative ES.

Integrated Outcomes Across Meta-Analyses

This section of Tables 2 to 5 presents the integrated outcomes across meta-analyses after the meta-analytic procedures that were conducted. Key outcomes include the integrated Hedges' *g*s, the number of individual meta-analyses contributing to integrated Hedges' *g*s, associated *SE*s, 95% CIs, *p* values, and results of heterogeneity analyses that assessed whether Hedges' *g*s of the original meta-analyses varied significantly. Key statistics include between-meta-analysis variance (Q_b), degrees of freedom, *p* values, and the ratio of between-meta-analysis variance to total variance (I^2). Considering the small sample size and potential low power of the analysis (Higgins & Green, 2011), we used a significance level

TABLE 3

Effects of within-class ability grouping on academic achievement

Outcomes of the individual original meta-analyses								
Meta-analysis	Student group	No. of studies (ESs)	Original mean ES[a]	Hedges' *g*[b]	Standard error[c]	95% CI	*p*	Narratives/interpretations in the original meta-analysis
Kulik and Kulik (1987)[1]	Overall (i.e., programs designed for all students)[d]	15	0.17	0.17	0.09	[−0.01, 0.34]	.06	"In the 15 studies, the average overall effect was 0.17 standard deviation. This average effect was different from zero at only a marginal level of significance, $t(14) = 2.10$, $p < .10$." (p. 27)
	High ability	6	0.29	0.27	0.08	[0.10, 0.43]	<.01	"Effects were similar for students of high, middle, and low ability taught in grouped and ungrouped classrooms; the respective average effect sizes were 0.29, 0.17, and 0.21." (p. 27)
	Medium ability	6	0.17	0.16	0.08	[−0.01, 0.32]	.06	
	Low ability	6	0.21	0.19	0.08	[0.03, 0.36]	.02	
Slavin (1987)	Overall	8	0.34[2]	0.32	0.09	[0.15, 0.49]	<.01	"Every study of within-class ability grouping in mathematics favored the practice, though not always significantly. The median effect size for the five randomized studies is 0.32; including matched studies makes the median only slightly higher." (p. 317)
	High ability	5	0.37[3]	0.33	0.08	[0.17, 0.49]	<.01	"Every subgroup gained more in classes using within-class ability grouping than in control (ungrouped) treatments." (p. 319)
	Medium ability	5	0.29[4]	0.26	0.08	[0.10, 0.42]	<.01	
	Low ability	5	0.53[5]	0.48	0.08	[0.32, 0.64]	<.01	
Kulik and Kulik (1992)	Overall	11	0.25	0.24	0.09	[0.07, 0.41]	.01	"The average overall effect of grouping in the 11 studies was to raise examination scores by 0.25 standard deviations, a significant but small effect." (p. 75)
	High ability	6	0.30	0.28	0.08	[0.11, 0.44]	<.01	"The average effect size was 0.30 for the higher ability students; 0.18 for the middle ability students; and 0.16 for the low-ability students." (p. 75)
	Medium ability	6	0.18	0.17	0.08	[0.00, 0.33]	.05	
	Low ability	6	0.16	0.15	0.08	[−0.02, 0.31]	.08	

(continued)

TABLE 3 (continued)

Outcomes of the individual original meta-analyses								
Meta-analysis	Student group	No. of studies (ESs)	Original mean ES[a]	Hedges' g^b	Standard error[c]	95% CI	p	Narratives/interpretations in the original meta-analysis
Mosteller et al. (1996)	Overall	3	0.40[6]	0.32	0.07	[0.18, 0.46]	<.01	"Among the three experiments involving within-class skill grouping presented here, two show considerable promise for within-class skill grouping; the other is neutral. " (p. 810)
Lou et al. (1996)[7]	Overall	51 (103 ESs)	0.17	0.17	0.09	[−0.01, 0.34]	.06	"On average, students learning in small groups within classrooms achieved significantly more than students not learning in small groups." (p. 439)
	High ability	≤18 (18 ESs)	0.28	0.27	0.09	[0.10, 0.45]	<.01	"While low-ability students, medium-ability students, and high-ability students all benefited from being placed in small groups (d = +0.37, +0.19, and +0.28), the effects were not uniform. In particular, low-ability students achieved significantly more than medium-ability students." (pp. 443–444)
	Medium ability	≤11 (11 ESs)	0.19	0.18	0.09	[0.01, 0.35]	.03	
	Low ability	≤24 (24 ESs)	0.37	0.36	0.09	[0.19, 0.54]	<.01	

Integrated outcomes across meta-analyses									
	Meta-analytic statistics					Heterogeneity across meta-analyses			
Grouping type/ student group	No. of meta-analysis	Integrated Hedges' g^e	SE	95% CI	p	$Q_b{}^f$	df	p	I^{2g}
Within-class grouping									
Overall	5	0.25	0.04	[0.18, 0.32]	<.01	3.42	4	.49	0.00
High ability	4	0.29	0.04	[0.11, 0.44]	<.01	0.43	3	.93	0.00
Medium ability	4	0.19	0.04	[0.11, 0.27]	<.01	1.03	3	.79	0.00
Low ability	4	0.30	0.08	[0.14, 0.45]	<.01	10.34	3	.02	70.99

Note. CI = confidence interval; ES = effect size; *df* = degrees of freedom; *SE* = standard error. Notations 1 to 6 describe how particular original mean ESs were computed or extracted from the original meta-analyses. They are included the Supplementary Appendix (available in the online version of the journal). Notations a to g are the same as in Table 2.

TABLE 4

Effects of cross-grade subject grouping on academic achievement

Outcomes of the individual original meta-analyses

Meta-analysis	Student group	No. of studies (ESs)	Original mean ES[a]	Hedges' *g*[b]	Standard error[c]	95% CI	*p*	Narratives/interpretations in the original meta-analysis
Kulik and Kulik (1987)	Overall	16	0.23	0.22	0.09	[0.05, 0.40]	.01	"The average effect size in the 16 studies was 0.23. This effect was different from zero at a marginal level of significance, $t(15) = 2.03$, $p < .10$." (p. 25)
	High ability	4	0.38[1]	0.33	0.08	[0.18, 0.48]	<0.01	Not given
	Medium ability	3	0.42[2]	0.34	0.07	[0.19, 0.48]	<0.01	
	Low ability	3	0.29[3]	0.23	0.07	[0.09, 0.37]	<0.01	
Slavin (1987)	Overall	14	0.40[4]	0.39	0.09	[0.22, 0.56]	<0.01	"Overall, the evidence in Table 3 consistently supports the use of the Joplin Plan. Joplin classes achieved more than control classes in 11 of 14 comparisons, with the remaining three studies finding no difference." (p. 311)
Kulik and Kulik (1992)	Overall	14	0.30	0.29	0.09	[0.12, 0.46]	<0.01	"The average effect size in the 14 studies was 0.30, a small effect but one that is significantly greater than zero." (p. 75)
Mosteller et al. (1996)	Overall	2	0.33	0.19	0.05	[0.09, 0.29]	<0.01	"Overall, the Joplin Plan treatment led to reading improvement with an effect size of 0.33. This effect size is larger than those of most of the XYZ skill-grouping studies reviewed earlier." (p. 808)

Integrated outcomes across meta-analyses

	Meta-analytic statistics					Heterogeneity across meta-analyses			
Grouping type/student group	No. of meta-analysis	Integrated Hedges' *g* [e]	Standard error	95% CI	*p*	Q_b[f]	*df*	*p*	I^2[g]
Cross-grade subject grouping[d]									
Overall	4	0.26	0.04	[0.17, 0.34]	<.01	4.23	3	0.24	29.10

Note. CI = confidence interval; ES = effect size; *df* = degrees of freedom; *SE* = standard error. Notations 1 to 4 described how particular original mean ESs were computed or extracted from the original meta-analyses. They are included in Supplementary Appendix (available in the online version of the journal). Notations a to c and e to g are the same as those of Table 2.

[d] The integrated Hedges' *g* was not reported by students of different achievement levels because there was only one meta-analysis (i.e., Kulik & Kulik, 1987) integrated the effects of cross-grade subject grouping by high-, medium-, and low-ability students.

TABLE 5

Effects of special grouping for the gifted on academic achievement

Outcomes of the individual original meta-analyses								
Meta-analysis	Student group	No. of studies (ESs)	Original mean ES[a]	Hedges' *g*[b]	Standard error[c]	95% CI	*p*	Narratives/interpretations in the original meta-analysis
Kulik and Kulik (1982)	Gifted overall	14	0.33	0.32	0.09	[0.15, 0.49]	<.01	
Kulik and Kulik (1984a)	Gifted overall	9	0.49	0.47	0.09	[0.30, 0.63]	<.01	"In studies of programs designed specifically for gifted and talented students, grouping raised achievement test scores by .49 standard deviations." (p. 5)
Kulik (1985)	Gifted overall	25	0.33	0.32	0.09	[0.15, 0.50]	<.01	"The average effects in the 25 studies of honors programs was to raise student examination scores by one third of a standard deviation. This is small to moderate effect. It is equivalent in the elementary school years to an increase of about three months on a grade equivalent scale."(p. 5)
Kulik and Kulik (1987)	Gifted overall	25	0.33	0.32	0.09	[0.15, 0.50]	<.01	"An effect size of 0.33 standard deviations means that teaching talented students in homogeneous rather than heterogeneous classes raised their scores on achievement tests by 0.33 standard deviations." (p. 25)
Goldring (1990)	Gifted overall	18	0.36	0.35	0.09	[0.18, 0.52]	<.01	"On the achievement outcomes, the average effect sizes generally favored the gifted students in the special classrooms over the gifted students in the regular classrooms." (p. 319)
Kulik and Kulik (1992)	Gifted overall	25	0.41	0.40	0.09	[0.23, 0.58]	<.01	"The average effect in the 25 studies was 0.41. This effect is moderate in size and significantly greater than an effect size of zero." (p. 76)

Integrated outcomes across meta-analyses									
	Meta-analytic statistics					Heterogeneity across meta-analyses			
Grouping type/ student group	No. of meta-analysis	Integrated Hedges' *g*[e]	Standard error	95% CI	*p*	Q_b[f]	*df*	*p*	I^2[g]
Special grouping for the gifted[d]									
Gifted overall	6	0.37	0.04	[0.30, 0.44]	<.01	2.29	5	0.81	.00

Note. CI = confidence interval; ES = effect size; *df* = degrees of freedom; *SE* = standard error. Notations a to c and e to g are the same as in Table 2.

[d] The integrated Hedges' *g* was not reported by students of different achievement levels because related meta-analyses focused on gifted students.

of .10 (as opposed to the conventional level of .05) when determining statistical significance for our heterogeneity analyses. The results of heterogeneity analyses should be interpreted with caution, given that Hedges' *g*s were converted from the original mean ESs based on the average of available *SE*s.

Between-Class Ability Grouping

A total of 36 Hedges' *g*s were computed from corresponding original mean ESs extracted from 11 meta-analyses. The values ranged from −0.54 to 0.19, 15 of them were negative, 2 equaled zero, and 18 were positive. The effects of between-class grouping on overall students' academic achievement was not statistically significant and there was considerable heterogeneity among the 11 mean ESs that contributed to the integrated Hedges' *g* for students overall (see Table 2).

For high-ability students, the integrated ES was small and not statistically significant, and the seven mean ESs associated with high-ability students appeared to be homogeneous. For medium-ability students, the integrated ES was small and not statistically significant, and there was significant heterogeneity among the associated 10 mean ESs. Similarly, for low-ability students, the integrated ES was also not statistically significant, and the eight mean ESs associated with low-ability students appeared to be homogeneous (see Table 2). Furthermore, the integrated ESs corresponding to the overall, high-, medium-, and low-ability students did not vary significantly. Taken together, integrated outcomes of the 11 ability grouping meta-analyses showed that the effects of between-class ability grouping on K–12 students' academic achievement was negligible, regardless of students' initial achievement levels or learning abilities.

Within-Class Ability Grouping

A total of 17 Hedges' *g*s were converted from the original mean ESs from five meta-analyses. These 17 Hedges' *g*s ranged from 0.15 to 0.48. The integrated outcomes showed that the effects of within-class grouping on overall students' academic achievement was positive and statistically significant (see Table 3). The mean ESs associated with all students, high-ability students, and medium-ability students appeared to be homogeneous. For high-ability students, the integrated ES was positive, small to moderate, and statistically significant. For medium-ability students, the integrated ES was positive, relatively small but statistically significant. For low-ability students, the integrated ES was moderate and statistically significant, and appeared to be heterogeneous. Furthermore, the integrated ESs for the all, high-, medium-, and low-ability students did not vary significantly. In sum, integrated outcomes of the five related meta-analyses showed that within-class ability grouping had at least small, positive, and significant impact on K–12 students' academic achievement, regardless their initial achievement or ability levels.

Cross-Grade Subject Grouping

A total of seven Hedges' *g*s were computed from the original mean ESs extracted from four meta-analyses. These seven ESs ranged 0.19 to 0.39. As can be seen in Table 4, the integrated outcomes showed that the effects of cross-grade subject grouping on overall students' academic achievement was positive, small, and statistically significant. There was minor heterogeneity among the four mean ESs that were associated with the overall students. Only one meta-analysis (i.e., Kulik & Kulik, 1987) aggregated the effects of cross-grade subject grouping by high-, medium-, and low-ability students.

Special Grouping for the Gifted

Six Hedges' *g*s were converted from the original mean ESs of six meta-analyses, ranging from 0.32 to 0.47. The integrated outcomes showed that the effects of special grouping for the gifted on gifted students' academic achievement overall was positive, moderate, and statistically significant (see Table 5). The six contributing ESs appeared to be homogeneous, suggesting that gifted students benefited from being placed in special groups or programs that were specifically designed to serve those with initial high achievement levels or learning potential.

Publication Bias

There was a small degree of publication bias among respective meta-analyses of between-class grouping and cross-grade subject grouping, which led to a slight overestimate of the mean ESs. For example, two meta-analyses were projected missing from the left side of the funnel plot, suggesting a presence of publication bias. After the trim-and-fill procedures, the mean ES decreased to 0.03, suggesting that the mean ES of 0.06 was a slight overestimate of the true effects of between-class ability grouping. There was no evidence of publication bias among meta-analyses of within-class grouping and special grouping for the gifted.

Methodological Quality of Meta-Analyses and Their Link to Outcomes

Table 6 shows the outcomes of comparing the integrated Hedges' *g*s of meta-analyses of low and those of moderate methodological quality. In the 11 meta-analyses of between-class ability grouping, the mean ESs of those of moderate methodological quality appeared to be smaller than those of low quality, but none of the differences was statistically significant. In the five meta-analyses of within-class grouping, the mean ESs of those of moderate quality appeared to be larger than those of low quality, but the difference was statistically significant only in the case of low-ability students. Furthermore, in all meta-analyses of cross-grade subject grouping or special grouping for the gifted, there were no significant differences in the mean ESs linking to differential methodological quality. Taken together, the effects of ability grouping did not manifest differently in meta-analyses of various methodological quality.

Mini Meta-Analysis of 12 Randomized Studies

Descriptive Information

Of the 12 studies included in the mini meta-analysis, 6 (50%) were published in peer-reviewed journals, 5 (41.6%) were doctoral dissertations, and 1 was a research monograph (i.e., Drews, 1963). Table 7 presents the key features of these 12 randomized studies, including the meta-analytic outcomes at the bottom. Five studies assessed the effects of between-class grouping, five assessed within-class grouping, and two assessed cross-grade subject grouping. Between-class grouping was mostly implemented in middle school to junior high students (i.e., Grades

TABLE 6

Comparisons of meta-analyses of low and moderate methodological quality

Grouping type/ student group	No. of meta-analyses		Integrated Hedges' *g*		Heterogeneity across low- and moderate-quality meta-analyses		
	Low	Moderate	Low	Moderate	Q_b	*df*	*p*
Between-class grouping							
Overall	7	4	0.03	−0.15	1.43	1	.23
High ability	4	3	0.07	0.02	0.68	1	.41
Medium ability	7	3	−0.03	−0.06	0.11	1	.74
Low ability	5	3	0.06	−0.04	2.43	1	.12
Within-class grouping							
Overall	2	3	0.20	0.28	0.93	1	.34
High ability	2	2	0.27	0.31	0.17	1	.68
Medium ability	2	2	0.16	0.23	0.58	1	.45
Low ability	2	2	0.17	0.43	9.27	1	.002
Cross-grade subject grouping							
Overall	2	2	0.26	0.28	0.02	1	.89
High ability	1		0.33		0.00	1	1
Medium ability	1		0.34		0.00	1	1
Low ability	1		0.23		0.00	1	1
Special grouping for the gifted							
Gifted overall	5	1	0.37	0.35	0.03	1	.86

Note. df = degrees of freedom. Low = low methodological quality; Moderate = moderate methodological quality.

7–9). Within-class grouping and cross-grade subject grouping were mostly employed in elementary students (i.e., Grades 1–6). Ability grouping was implemented for instruction of math in four studies, for arithmetic in three studies, and for English and reading in three studies. The duration of ability grouping in the studies ranged from 16 weeks to 3 years, and the number of participants ranged from 52 to 480. The 12 studies involved approximately 2,434 students of Grades 1 to 9 in the United States. The majority of studies compared ability-grouped students with their peers who received traditional instruction in heterogeneous classes.

ESs were extracted corresponding to the number of comparisons in a study. For example, as Table 7 shows, two ESs were extracted from Slavin and Karweit (1985, Experiment 1), one for the comparison between within-class grouping and Missouri Mathematics Program (a modified instruction program at the time) and the other for the comparison between within-class grouping and individualized instruction that was similar to cooperative learning. ESs were also extracted corresponding to students' initial ability levels as applicable. For instance, although there was only one comparison in the Dewar (1963) study (i.e., within-class grouping vs. traditional whole-class instruction), three ESs were extracted (for

TABLE 7

Meta-analysis of 12 randomized experimental studies of ability grouping

Outcomes of the individual experimental studies

Study	Subject	Participants	Duration	Experimental group	Comparison group	Hedges’ *g*	*SE*	95% CI	*p*
Between-class grouping									
Ford (1974)	Math	100 9th graders in an urban high school in New York City	One semester (term)	Between-class grouping (high-ability and low-ability students were grouped into different classes)	Traditional heterogeneous class				
				High ability		0.50	0.40	[−0.21, 1.29]	.22
				Low ability		0.45	0.37	[−0.28, 1.19]	.22
Bicak (1964)	Science (meteorology)	75 Eighth graders at the University of Minnesota High School	One quarter (semester)	Between-class grouping (i.e., high ability)	Traditional heterogeneous group (Note: only high-ability students in this group were compared to the ability-grouped high-ability students)	0.31	0.30	[−0.28, 0.91]	.30
				Between-class grouping (i.e., low ability)	Traditional heterogeneous group (Note: only low-ability students in this group were compared with the ability-grouped low-ability students)	0.12	0.29	[−0.46, 0.70]	.69
Drews (1963)	English	432 9th graders in the Lansing Public Schools	1 School year	Homogeneous grouping (i.e., three classes for 78 superior, 114 average, and 59 slow students)	Heterogeneous grouping (i.e., 23 superior students, 137 average students, and 21 slow students were distributed to three heterogeneous groups)				
				High ability		0.17	0.16	[−0.16, 0.49]	.31
				Average ability		0.08	0.09	[−0.10, 0.25]	.37
				Low ability		0.14	0.18	[−0.21, 0.48]	.44

(continued)

TABLE 7 (continued)

Outcomes of the individual experimental studies

Study	Subject	Participants	Duration	Experimental group	Comparison group	Hedges' *g*	*SE*	95% CI	*p*
Fick (1962)	7th grade core classes	168 7th graders in the Olathe Junior High School	1 School year	Between-class grouping (high, average, low)	Heterogeneous classes (students of high-, average-, and low-ability students were distributed to three heterogeneous classes)	0.10	0.11	[−0.11, 0.32]	.33
Mikkelson (1962)	Math	140 (70 7th graders selected from the top 20% of the 7th grade; 70 8th graders selected from the top 20% of the 8th grade)	2 school years (i.e., 7th and 8th grade)	Between-class grouping (selected top 35 7th graders were placed together in one homogenous class) In 7th grade, both ability-grouped class and control classes received the same curriculum	Traditional heterogeneous class (selected top 35 7th graders were distributed into other four control classes of the grade)	0.24	0.19	[−0.13, 0.62]	.20
				Between-class grouping (selected top 35 8th graders were placed together in one homogenous class) In 8th grade, the ability-grouped homogeneous class received acceleration in math curriculum	Traditional heterogeneous class (selected top 35 8th graders were distributed among four control classes of the grade)	0.15	0.19	[−0.23, 0.53]	.44
Within-class grouping									
Slavin and Karweit (1985; Experiment 1)	Math	345 4th to 6th graders in 15 classes in Delaware	18 Weeks	Within-class grouping (i.e., Ability-grouped Active Teaching [AGAT], Students in each AGAT class were divided into two groups: high group—about 60% of the students; low group—40%)	Missouri Mathematics Program (MMP) (i.e., a modified instruction program at the time)	0.44	0.36	[−0.27, 1.14]	.22
					Individualized instruction (i.e., Team Assisted Individualization [TAI], similar to cooperative learning)	0.04	0.09	[−0.14, 0.21]	.67

(continued)

TABLE 7 (continued)

Outcomes of the individual experimental studies

Study	Subject	Participants	Duration	Experimental group	Comparison group	Hedges' *g*	*SE*	95% CI	*p*
Slavin and Karweit (1985; Experiment 2)	Math	480 3rd to 5th graders 22 classrooms in Maryland	16 weeks	Within-class grouping (i.e., AGAT)	MMP	0.59	0.09	[0.41, 0.77]	.00
					Individualized instruction (i.e., TAI, similar to cooperative learning)	0.38	0.18	[0.03, 0.73]	.03
					Traditional whole-class instruction	0.75	0.10	[0.55, 0.95]	.00
Dewar (1963)	Arithmetic	8 6th Grade classrooms in a suburban, upper middle class in Kansas	1 school year	Within-class grouping	Traditional whole-class instruction (each class was divided into 3 groups based on the pretest results, for statistical purposes, but NOT for instruction purposes)				
				High ability (i.e., Group 1)		0.54	0.26	[0.04, 1.05]	.03
				Medium ability (i.e., Group 2)		0.42	0.23	[−0.02, 0.87]	.06
				Low ability (i.e., Group 3)		0.66	0.26	[0.14, 1.18]	.01
Smith (1961)	Arithmetic	190 2nd through 5th graders in the White elementary schools of the City of Lake Charles	One semester	Within-class grouping, i.e., experimental classes spent at least 75% of the daily arithmetic period working in intra class groups	Traditional class ("were taught by the 'class-as-a-whole' procedure")				
				High ability (Group A)		0.22	0.13	[−0.03, 0.47]	.09
				Average ability (Group B)		0.25	0.13	[−0.01, 0.50]	.06
				Low ability (Group C)		0.41	0.12	[0.18, 0.64]	.00
Wallen and Vowles (1960)	Arithmetic	112 6th graders in 4 arithmetic classes in Salt Lake City	1 school year	Within-class grouping (In all 4 classes, the teachers used the ranking of CAT scores to divide the class into four homogeneous groups of 5 to 8 students)	Traditional whole-class instruction (i.e., nongrouping teaching)	0.07	0.13	[−0.19, 0.33]	.58

(continued)

TABLE 7 (continued)

Outcomes of the individual experimental studies

Study	Subject	Participants	Duration	Experimental group	Comparison group	Hedges' *g*	*SE*	95% CI	*p*
Cross-grade subject grouping									
Morgan and Stucker (1960)	Reading	90 Matched pairs of 5th and 6th graders in southeastern Michigan	1 school year	Subject ability grouping (i.e., the Joplin Plan)	Traditional whole-class instruction				
				Fast achievers	Fast achievers	0.43	0.23	[−0.02, 0.87]	.06
				Slow achievers	Slow achievers	0.89	0.31	[0.29, 1.50]	.00
Jones et al. (1967)	Reading	52 first graders	1.5 and 3 years	Subject ability grouping (i.e., nongraded reading program)	Within-class grouping (i.e., in 1st, 2nd, and 3rd grades, students were grouped in three levels)				
				1.5 years	1.5 years	0.71	0.16	[0.39, 1.03]	.00
				3 years	3 years	0.33	0.16	[0.01, 0.65]	.05

Meta-analytic outcomes

	Meta-analytic statistics					Heterogeneity across studies			
Grouping type	No. of studies	Mean Hedges' *g*	*SE*	95% CI	*p*	Q_b	*df*	*p*	I^2
Between-class grouping	5	0.15	0.07	[0.01, 0.29]	.03	1.07	4	.90	0.00
Within-class grouping	5	0.33	0.09	[0.14, 0.52]	<.01	11.76	4	.02	65.98
Cross-grade subject grouping	2	0.57	0.13	[0.32, 0.83]	<.01	0.33	1	.56	0.00

Note. CI = confidence interval; ES = effect size; *df* = degrees of freedom; *SE* = standard error.

high-, medium-, and low-ability students). Similarly, two ESs were extracted from the Morgan and Stucker (1960) study, one for the fast achievers and the other for slow achievers.

The Jones et al. (1967) study was unique. Jones et al. compared the effects of cross-grade subject grouping (in a nongraded reading program) of first, second, and third graders with that of within-class grouping in which students in each grade level (i.e., Grades 1–3) grouped into high, medium, and low reading levels. Student learning outcomes were collected after 1.5 years and again after the 3 years. Therefore, two ESs were extracted from the study, one corresponding to the 1.5 years and the other to the 3 years of grouping duration. A total of 26 ESs, ranging from 0.04 to 0.89, were extracted from the 12 studies. Separate meta-analytic integrations were conducted on the five studies of between-class grouping, five studies of within-class grouping, and on the two studies of cross-grade subject grouping. Analyses were conducted with mixed-effects models and the individual studies as the unit of analysis in the CMA software.

Meta-Analytic Outcomes

The effects of between-class grouping on middle school and junior high students' academic achievement were positive, small, and statistically significant and appeared to be homogeneous across the five studies. The effects of within-class grouping on elementary students' academic achievement were positive, moderate, and statistically significant, but they varied significantly across the five studies. The effect of cross-grade subject grouping on elementary students' academic achievement was also positive, moderate, and statistically significant and was similar in the two studies. Furthermore, the effects of between-class grouping and cross-grade subject grouping were significantly different from each other, but there was no significant differences between all other pair-wise comparisons, and these results are not reported in Table 7. In sum, findings of these 12 randomized experimental studies showed that students obtained small to moderate benefits from between-class grouping, within-class grouping, and especially cross-grade subject grouping.

Results on Acceleration

Descriptive Information

A Brief Overview of Six Acceleration Meta-Analyses

Six acceleration meta-analyses reviewed at least 125 unique primary studies that involved at least 75,582 participants. Primary studies were published as early as 1918 up through 2008. Four primary studies were published in the 1920s, 7 in the 1930s, 1 in the 1940s, 11 in the 1950s, 35 in the 1960s, 12 in the 1970s, 25 in the 1980s, 17 in the 1990s, and 13 in the 2000s.

The Kulik and Kulik (1984b) meta-analysis was the earliest meta-analytic review on acceleration. They examined 26 studies of elementary and secondary acceleration that focused on grade skipping, curricular compacting, and adding summer session to school. Results indicated that the academic performance of accelerated students exceeded the performance of same age and intelligence students who were not accelerated by almost one academic year. Moreover, no performance differences were found when accelerated students were compared with older nonaccelerated peers. Rogers's (1991) dissertation expanded the synthesis of acceleration work to include many additional forms of acceleration such as early entrance to school, Advanced Placement courses, concurrent enrollment, and mentorship. Rogers (1991) concluded that "educational decision-makers have been offered a fairly well research-supported menu of accelerative options that results in significant academic achievement gains" (p. 208).

Kulik (2004) reanalyzed studies that had previously been assessed by Kulik and Kulik (1984b), Rogers (1991), and Kent (1992), who primarily focused on social and emotional outcomes of acceleration. Specifically seeking to update her previous findings, Rogers (2008) meta-analyzed studies that had been conducted after her 1991 study, and concluded, "In general there is a powerful academic effect to be gained from engaging in a variety of forms of acceleration" (p. 2). In a meta-analysis focusing on comprehensive school reform, Borman, Hewes, Overman, and Brown (2003, p. 162) included an analysis of six studies of accelerated schools that they claimed provided "promising evidence for effectiveness," but concluded that there were too few studies to make confident general conclusions about effectiveness. The most recent meta-analysis on acceleration by Steenbergen-Hu and Moon (2011) sought to build on the original acceleration meta-analysis by analyzing all acceleration research conducted after Kulik and Kulik's (1984b) meta-analysis. Similar to the other meta-analyses on acceleration, Steenbergen-Hu and Moon (2011, p. 39) concluded that "acceleration influences high-ability learners in positive ways."

Study Overlap Across Meta-Analyses

As reported in Supplementary Table S3 (available in the online version of the journal), the number of primary studies reviewed in the six acceleration meta-analyses ranged from 6 (Borman et al., 2003) to 81 (Rogers, 1991). Using the 81 studies in Rogers's (1991) meta-analysis as the reference group, Kulik and Kulik (1984b) and Kulik (2004) had high degrees of study overlap, 100% and 88.9%, respectively. The meta-analyses by Borman et al. (2003) and Steenbergen-Hu and Moon (2011) had low degrees of study overlap (0% and 7.1%, respectively) with Rogers's (1991) meta-analysis. Rogers's (2008) meta-analysis did not provide references for included primary studies. However, because it only included studies from 1990 to 2008 and was conducted by the same author, it likely had no overlap with Rogers (1991). Again, varied inclusion criteria and the number of included studies in meta-analyses might explain the low degree of study overlap in Borman et al.'s (2003), Rogers's (2008), and Steenbergen-Hu and Moon's (2011) meta-analyses.

Ways That Acceleration Meta-Analyses Aggregated Outcomes

The acceleration meta-analyses aggregated the outcomes of primary studies and reported ES in two different ways (see Table 8). Three reported ESs primarily by the type of comparison groups—whether the accelerated students were compared with their same-age, older, mixed-age peers, or all types of comparison groups combined (see Notation 3 of Table 9 for detailed explanation; Kulik, 2004; Kulik & Kulik, 1984b; Steenbergen-Hu & Moon, 2011). The outcomes of these meta-analyses showed the academic effects of acceleration after taking age into

TABLE 8

Initial outcomes extracted from acceleration meta-analyses

Meta-analysis	Comparison group	Form of acceleration	No. of studies (ESs)	Original mean ES^{a}	Related statistics	Narratives/interpretations of the original meta-analyses
Kulik and Kulik (1984b)	Same-age peers (elementary)		3	1.43	*SE* = 0.76	"Talented students are able to handle the academic challenge that accelerated programs provide . . . talented youngsters who were accelerated into higher grades performed as well as the talented, older pupils already in those grades. . . . in the subjects in which they were accelerated, talented accelerates showed almost a year's advancement over talented same-age nonaccelerates." (p. 421)
	Same-age peers (junior high)		9	0.76	*SE* = 0.15	
	Same-age peers (senior high)		1	0.28	—	
	Older peers (elementary)		5	0.16	*SE* = 0.17	
	Older peers (junior high)		8	−0.01	*SE* = 0.11	
Rogers (1991)		Early entrance	10 (31)	0.49		"In general, significant academic effects (ES > .30) were found for all but three options: concurrent enrollment, Advanced Placement, and Combined options." (p. 162) "Educational decision-makers have been offered a fairly well research-supported menu of accelerative options that result in significant academic achievement gains." (p. 208)
		Grade skipping	9 (25)	0.46^{b}		
		Nongraded classes	3 (7)	0.43		
		Compacting	4 (9)	0.83^{c}		
		Telescoping	11 (53)	0.45		
		Concurrent enrollment	8 (11)	0.22^{d}		
		Subject acceleration	10 (28)	0.59^{e}		
		AP	6 (13)	0.27		
		Mentorship	5 (21)	0.47		
		Credit by exam	5 (28)	0.59		
		Early admission to college	8 (28)	0.30^{f}		
		Combined options	2 (18)	0.13		
Borman et al. (2003)	Elementary, all studies	Accelerated school	6	0.09	95% CI [0.05, 0.12], *SE* = 0.09	Categorized as "Promising evidence of effectiveness"
	Elementary, studies with a comparison group	Accelerated school	3	0.21	95% CI [0.07, 0.36], *SE* = 0.07	Categorized as "Promising evidence of effectiveness"
	Elementary, evaluation conducted by a third party	Accelerated school	2	0.13	95% CI [−0.06, 0.31], *SE* = 0.10	Categorized as "Promising evidence of effectiveness"
Kulik (2004)^{g}	Same-age peers (elementary)		3	1.62		"The overall message from these studies is therefore unequivocal: Acceleration contributes greatly to the academic achievement of bright students." (p. 15)
	Same-age peers (junior high)		6	0.73		
	Same-age peers (senior high)		2	0.47		"The accelerated students did just as well as the bright students in the grades into which they moved." (p. 18)
	Same-age peers (total)		11	0.82		

(continued)

TABLE 8 (continued)

Meta-analysis	Comparison group	Form of acceleration	No. of studies (ESs)	Original mean ES[a]	Related statistics	Narratives/interpretations of the original meta-analyses
	Older peers (elementary)		5	−0.02		"The meta-analytic results show that bright students almost always benefit from accelerated programs of instruction. Two major findings support this conclusion. First, on achievement tests, bright accelerated youngsters usually perform like their bright, older non-accelerated classmates. Second, the accelerated youngsters usually score almost one grade-level higher on achievement tests than bright, same-age nonaccelerated students do." (p. 20)
	Older peers (junior high)		6	−0.17		
	Older peers (senior high)		2	0.11		
	Older peers (total)		14	0		
Rogers (2008)		Accelerated honors/special school	2	1.04		". . . in general there is a powerful academic effect to be gained from engaging in a variety of forms of acceleration, either grade-based or subject-based and either as an in-school option or an out-of-school option." (p. 2)
		AP	7 (12)	0.62		
		Compacting	1 (18)	0.2		
		Computer/online courses	5 (20)	0.74		
		Early college admission	11 (20)	0.25		
		Early entrance to school	5 (8)	0.3		
		Grade skipping	4	0.37		
		Individualized Curriculum	1	2.35		
		IB program	2 (6)	0.54		
		Mentorship	4 (9)	0.22		
		Sat. classes on university campus	1	1.56		
		Subject acceleration	9 (34)	0.48		
		Summer university courses	11 (19)	0.45		
		Talent search programs	6 (21)	0.34		
Steenbergen-Hu and Moon (2011)	Overall		28	0.18	95% CI [−0.072, 0.431], *SE* = 0.128	". . . the findings from this meta-analysis generally confirm the positive influence of acceleration on high-ability learners." (p. 51)
	Same-age peers		13	0.396	95% CI [0.029, 0.762], *SE* = 0.187	
	Older peers		9	0.224	95% CI [−0.212, 0.660], *SE* = 0.222	
	mixed-age peers		6	−0.323	95% CI [−0.842, 0.197], *SE* = 0.265	

Note. CI = confidence interval; ES = effect size; *df* = degrees of freedom; *SE* = standard error.

[a] All original mean effects sizes are Cohen's *d*, with the exception of Steenbergen-Hu and Moon (2011), which reported Hedges' *g*s.

[b] In Rogers's (1991) meta-analysis, the mean ES of 0.46 was chosen over the initial one of 0.56 because the former was the result of removing "possible outlier."

[c] Similarly, the mean ES of 0.83 was chosen over the initial one of 1.23 because the former was the result of removing two outliers.

[d] Similarly, the mean ES of 0.22 was chosen over the initial one of 0.21 because the former was the result of removing two outliers.

[e] Similarly, the mean ES of 0.59 was chosen over the initial one of 0.98 because the former was the result of removing nine outliers.

[f] Similarly, the mean ES of 0.30 was chosen over the initial one of 0.35 because the former was the result of removing possibly one outlier.

[g] The published meta-analysis only provided median ESs of results based on studies using same age and older control students. Relying on information provided in Tables 1 and 2 (same age and older control, respectively) of the meta-analysis, we calculated sample weighted mean ESs.

TABLE 9

Effects of acceleration on academic achievement

Outcomes of the individual original meta-analyses[a]							
Meta-analysis	Way of outcome aggregation[b]	No. of studies	Original mean ES	Hedges' g	Standard error	95% CI	p
Meta-analyses that reported effect sizes by the type of comparison groups							
Kulik and Kulik (1984b)	Same-age peers	13	0.82	0.79	0.22	[0.36, 1.23]	<.01
	Older peers	13	0.08	0.08	0.22	[−0.36, 0.51]	.73
Kulik (2004)	Same-age peers	11	0.91	0.88	0.22	[0.44, 1.31]	<.01
	Older peers	14	−0.02	−0.02	0.22	[0.46, 0.42]	.93
Steenbergen-Hu and Moon (2011)	Same-age peers	13	0.40	0.40	0.23	[−0.05, 0.85]	.08
	Older peers	9	0.22	0.22	0.23	[−0.23, 0.67]	.34
	Mixed-age peers	6	−0.32	−0.32	0.23	[−0.77, 0.13]	.16
	All types of comparison groups combined[c]	28	0.18	0.18	0.23	[−0.27, 0.63]	.43
Meta-analyses that reported effect sizes by the form of acceleration							
Rogers (1991)	All forms of acceleration combined[d]	81	0.47	0.47	0.23	[0.02, 0.92]	.04
Borman et al. (2003)	All forms of acceleration combined	6	0.14	0.13	0.21	[−0.29, 0.55]	.54
Rogers (2008)	All forms of acceleration combined	69	0.68	0.68	0.23	[0.23, 1.12]	<.01

TABLE 9 (continued)

Integrated outcomes across meta-analyses									
	Meta-analytic statistics					Heterogeneity across meta-analyses			
Way of outcome aggregation	No. of meta-analysis	Integrated Hedges' g	Standard error	95% CI	p	Q_b	df	p	I^2
Same-age peers	3	0.70	0.15	[0.41, 0.98]	<.01	2.51	2	.29	20.25
Older peers	3	0.09	0.13	[−0.16, 0.35]	.49	0.56	2	.75	0
All types of comparison groups combined	1	0.18	0.23	[−0.27, 0.63]	.43	0	0	1	0
All forms of acceleration combined	3	0.42	0.16	[0.10, 0.73]	.01	3.17	2	.21	36.86
Integrated overall[e]	6	0.35	0.08	[0.20, 0.51]	<.01	17.98	9	.04	49.95

Note. CI = confidence interval; ES = effect size; *df* = degrees of freedom; *SE* = standard error.

[a] Unlike the cases in ability grouping meta-analyses (see Tables 2–5), in this table the section of *Outcomes of the Individual Original Meta-Analyses* presents average ESs combined within each acceleration meta-analysis.

[b] *Way of Outcome Aggregation* denotes the way each original acceleration meta-analyses aggregated the outcomes of primary studies and reported ESs.

[c] In addition to aggregating study outcomes by same-age, older, and mixed-age peers, the Steenbergen-Hu and Moon (2011) meta-analysis also aggregated outcomes of all included studies and termed "Overall" as one of the ways of aggregation (see Table 8). We renamed the "Overall" as *all types of comparison groups combined* in this second-order meta-analysis to prevent possible confusions with the term "integrated overall" that we used to denote the overall aggregated outcomes of all six acceleration meta-analyses regardless of their ways of aggregating outcomes.

[d] *All forms of acceleration combined* is assigned as a *way of outcome aggregation* for the original meta-analyses that reported ESs by the form of acceleration. In each of these three meta-analyses, the *Original Mean ESs* was the average of all ESs associated with various forms of acceleration as shown in Table 8.

[e] *Integrated overall* denotes the overall aggregated outcomes of all six acceleration meta-analyses regardless of their ways of aggregating outcomes.

account. The other three reported ESs by specific form of acceleration, such as grade skipping or curriculum compacting (Borman et al., 2003; Rogers, 1991, 2008). Of the 46 ESs extracted, 42 (91.3%) were Cohen's *d*s or an equivalent and four were Hedges' *g*s.

Integrating Mean Effect Sizes Across Meta-Analyses

Outcomes of Individual Meta-Analyses

Table 9 presents average ESs for the acceleration meta-analyses with results based on the way each original meta-analysis aggregated the outcomes. For each of the three meta-analyses that reported ESs by the form of acceleration, ESs associated with all forms of acceleration were averaged together. For instance, averaging the 14 ESs associated with early entrance, grade skipping, nongraded classes in Rogers (2008) led to an original mean ES of 0.68. Taken together, 11 mean Hedges' *g*s represented the outcomes of the six original meta-analyses. Similar to the ability grouping meta-analyses, each original mean ES was converted to Hedges' *g*. The 11 Hedges' *g*s ranged from −0.32 to 0.88, nine of them were positive, two were negative, and four were statistically significant. The values of eight (73%) of the Hedges' *g*s were equal to their corresponding original ESs and three (27%) were slightly smaller. The average of the 11 Hedges' *g*s was identical to that of the 11 original mean ESs.

Integrated Outcomes Across Meta-Analyses

The integration of ESs across meta-analyses was conducted by each way of outcome aggregation: same-age peers, older peers, all types of comparison groups combined, and all forms of acceleration combined. Only one meta-analysis (Steenbergen-Hu & Moon, 2011) reported ESs by mixed-age peers and all types of comparison group combined. Heterogeneity analyses showed that the ESs associated with comparisons between accelerated students and their nonaccelerated mixed-age peers was statistically significantly different from the integrated ESs associated with other ways of aggregation. This ES was therefore excluded from further analyses. The remaining 10 ESs were combined to generate an integrated overall ES.

Effects of acceleration on K-12 students' academic achievement was positive, at least moderate in magnitude, and statistically significant when assessed by comparing accelerated students with their nonaccelerated same-age peers; the effects of acceleration as found in the three meta-analyses did not vary significantly. Accelerated students did not significantly outperform their nonaccelerated older peers. Accelerated students also did not significantly surpass their nonaccelerated peers when age was not taken into account, as shown in one meta-analysis that combined the overall outcomes of studies of acceleration Heterogeneity analyses showed that the effects of acceleration did not differ significantly by whether they were assessed through comparing accelerated students with their nonaccelerated same-age peers, older peers, or through a relatively vague manner that ignored possible influences of age.

The three meta-analyses that examined the effects by form of acceleration showed that acceleration appeared to have a positive, moderate, and statistically significant impact on students' academic achievement. Heterogeneity analyses showed that the effects of acceleration did not differ significantly by whether they were assessed taking into account of influences of an age factor or pending particular forms of acceleration. Overall, acceleration had a positive, near moderate, and statistically significant impact on students' academic achievement, although effects varied significantly across studies.

Methodological Quality of Meta-Analyses

Supplementary Table S4 (available in the online version of the journal) presents information on the methodological quality of the six acceleration meta-analyses assessed with AMSTAR (Shea et al., 2009). Four of the six meta-analyses were categorized as having moderate methodological quality and two were categorized as having low quality (Rogers, 1991, 2008). An example of methodological weakness is that only three of the meta-analyses conducted a comprehensive search of the literature and reported seeking unpublished studies (Borman et al., 2003; Kulik & Kulik, 1984b; Steenbergen-Hu & Moon, 2011). Steenbergen-Hu and Moon's meta-analysis was the only one among the six that assessed the possibility of publication bias. Moderator analyses found that the effects of acceleration shown in the two meta-analyses of low methodological quality appeared to be greater than those in the four meta-analyses of moderate quality, but the difference was not statistically significant.

Publication Bias

A small degree of publication bias was present, which led to a slight overestimate of the integrated ESs. Specifically, two meta-analyses were projected missing from the left side of the funnel plot; the integrated overall ES changed from 0.38 to 0.33 after the trim-and-fill procedures. This finding should be viewed as fairly tentative, given that only six acceleration meta-analyses were involved.

Discussion

Outcomes of 13 ability grouping meta-analyses collectively showed that students benefited, at least to a small degree, from within-class grouping, cross-grade subject grouping, and special grouping for the gifted, whereas the benefits were negligible from between-class grouping. Overall, high-, medium-, and low-ability students benefited equally from ability grouping. The effects of within-class and cross-grade subject grouping are especially noteworthy given that research has consistently shown benefits for these types of acceleration. Gifted students benefited greatly from being placed in special groups or programs that were specifically designed to serve them, although this finding was based on only six meta-analyses, five of which had low methodological quality. Overall, these findings provide support for using ability grouping to meet the learning needs of students. They also remind us of the necessity to examine the effects of ability grouping by specific type, which is often overlooked by the general public.

Our mini meta-analysis of 12 RCTs revealed that students had small to moderate benefits from between-class grouping, within-class grouping, and especially cross-grade subject grouping. Interestingly, these ESs appeared to be larger than those reported in previous meta-analyses in which the majority of included studies were conducted with designs less rigorous than experimental ones. As the current second-order meta-analysis of ability grouping showed, the integrated mean ESs for between-class grouping ranged from −0.04 to 0.06, those of within-class grouping ranged from 0.19 to 0.30, and that of cross-grade subject grouping was $g = 0.26$. However, these findings are tentative given that the mini meta-analysis was based on only 12 studies.

Most important, our mini meta-analysis serves as a focal lens to revisit relevant outcomes of previous meta-analyses that had reviewed a subset of the same randomized studies. In the case of between-class grouping, we found that previous meta-analyses generally underestimated its effects. Mosteller et al.'s (1996) meta-analysis is comparable with our mini meta-analysis because they reviewed 10 randomized or nearly randomized studies of between-class grouping. Mosteller et al. reported that the average ES across the 10 randomized studies was 0 and not statistically significant, whereas our mini meta-analysis showed that integrated mean ES across the five studies of between-class grouping was 0.15 and statistically significant (see Table 7). Moreover, we found similar results when comparing the specific ESs extracted from each of the four common studies that were reviewed by three meta-analyses, Mosteller et al. (1996), Slavin (1990), and the current mini-meta-analysis. We wonder whether an underestimation of between-class grouping's effects contributed to the opposition to ability grouping in the 1980s and 1990s (e.g., Oakes, 1985; Slavin, 1993).

Previous meta-analyses and our mini meta-analysis yielded similar findings regarding the effects of within-class grouping when only randomized studies are considered. Slavin's (1987) meta-analyses had three randomized studies of within-class grouping in common with the current meta-analysis (i.e., Dewar, 1963; Smith, 1961; Wallen & Vowles, 1960). The ESs of these three studies were 0.55, 0.69, and 0.07 in Slavin's (1987) meta-analysis, and relevant ESs ranged from 0.07 to 0.66 in the current mini meta-analysis. Our mini meta-analysis also yielded similar findings concerning the effects of cross-grade subject grouping as previous studies. Our mini meta-analysis and Slavin's (1987) both reviewed two randomized studies of cross-grade subject grouping (i.e., Jones et al., 1967; Morgan & Stucker, 1960). For the Jones et al. (1967) study, Slavin (1987) reported an ES of 0.32 for high-ability students and 0.94 for low-ability students, and the current meta-analysis found ESs of 0.43 and 0.89, respectively. For the Morgan and Stucker (1960) study, both Slavin (1987) and the current meta-analysis reported an ES of 0.33 after the grouping was implemented for 3 years.

All six acceleration meta-analyses reported positive effects despite aggregating study outcomes in two different ways. Three meta-analyses aggregated study outcomes by types of comparison groups (e.g., nonaccelerated same-age vs. older peers). These meta-analyses showed that accelerated students significantly outperformed their nonaccelerated same-age peers to at least a moderate degree. However, their performance was not significantly greater than those of their nonaccelerated older peers. The remaining three meta-analyses aggregated study outcomes based on specific forms of acceleration. Outcomes of these meta-analyses showed that acceleration appeared to have a positive, moderate, and statistically significant impact on students' academic achievement. Overall, results from all six meta-analyses suggest that acceleration has a positive, near moderate, and statistically significant impact on accelerated students' academic achievement.

The Paradox of Research and Implementation

Findings of the current study underscore the long-standing paradox between empirical support for ability grouping and acceleration and the lack of policy to support greater implementation in schools (Kulik & Kulik, 1984b). If such a long history of research shows the effectiveness of most types of ability grouping and acceleration, the question of why it is not more universally implemented looms large for educators, parents, and policy makers. Such questions are apt, especially given how eager we are as a society to find educational interventions that are effective and can be implemented on a large scale for relatively low costs. The ill-founded concerns about socialization mentioned in the introduction (e.g., Belfi et al., 2012; Oakes, 2008; Smith, 1945) likely plague the application of acceleration and ability grouping for individual students.

Moreover, education administrators may have perverse incentives to avoid acceleration. For example, although acceleration can often actually save schools money because students spend fewer years in school, it can also "cost" schools money. Because school funding is often allocated based on headcounts and accelerated students spend fewer years in school, schools receive fewer dollars overall, or in the case of dual enrollment, may have to spend some of those dollars outside the district. Similarly, in states that offer open enrollment, students could leave a district for one where their needs are better met. Moreover, in the age of accountability via test score performance, keeping students who could be accelerated with their same-age peers can boost average test scores, regardless of whether the students are learning. Optimistically, we hope that accumulating research evidence would help catalyze the development of better policy and end the paradox of empirical support paired without widespread implementation.

Limitations

Findings of the current two second-order meta-analyses need to be understood considering two limitations. First, the conversion of the original mean ESs to the Hedges' *g*s relied on the average of the available *SE*s. This approach is particularly problematic for issues concerning between-meta-analysis variances. This approach, however, was at least as good as the most common way that researchers currently conduct second-order meta-analyses of education or psychology research—computing the mean ESs across meta-analyses without considerations of sampling errors or variance between meta-analyses (Borenstein et al., 2009; Cooper & Koenka, 2012). Also, although some preliminary equations and statistical methods for computing sampling errors in second-order meta-analyses have been developed (Schmidt & Oh, 2013), their applications are often restricted by missing information from the first-order meta-analyses (as is the case for the current study).

Overall, the current second-order meta-analyses employed the most feasible approach given the data available. We also implemented extra procedures to ensure that this approach yielded fair estimates of the effects. For example, we compared Hedges' *g*s with corresponding original mean ESs from which they were converted using the average of the available *SE*s, and examined whether they changed substantially if calculated based on available alternative information such as confidence intervals. Results showed that Hedges' *g*s were relatively conservative but fair counterparts of the original mean ESs. This finding alleviates the concern to some degree.

The second limitation is that the current two second-order meta-analyses, despite their comprehensiveness, cannot replace a new first-order meta-analysis that synthesizes the most current research, especially on the academic effects of ability grouping. The 13 ability grouping meta-analyses reviewed research between 1922 and 1994, and the most recent meta-analyses that met the inclusion criteria of the current second-order meta-analysis were conducted almost two decades ago. Although research on ability grouping largely ground to a halt during the 1990s, a substantial body of new research has appeared in recent years as ability grouping regained favor in the early 2000s (Olszewski-Kubilius, 2013). The new body of research advances the field in several ways. For example, a number of studies that analyze national large-scale data sets have appeared since 1996 (e.g., Chmielewski, Dumont, & Trautwein, 2013). Findings of such studies may be more generalizable to a larger or different study population than many earlier studies. Furthermore, recent studies benefit from advanced econometric methods such as regression discontinuity designs, propensity score matching, and multilevel modeling (e.g., Ruhose & Schwerdt, 2015), which may provide more robust and reliable estimates of the effects of ability grouping. As such, this substantial body of new research on ability grouping warrants a new meta-analysis. Such an endeavor would likely overcome some previous limitations and advance current knowledge.

Future Research

There is likely a continued need for direct and conceptual replications of many primary studies because the nature of education has changed radically since many of the studies were conducted in terms of generation or cohort effects (Makel & Plucker, 2014). Future research might examine accelerative interventions in specific domains, at specific ages, and on students of diverse social–economic status (Plucker & Harris, 2015). For example, does domain-specific acceleration in math differ from domain-specific acceleration in science or language arts in the fourth grade versus the seventh grade, especially since domains seem to vary in the extent to which experience or maturity can affect the degree to which progress is accelerated? Moreover, meta-analyses need to be conducted on the relative effectiveness of each *type* of acceleration once there are sufficient studies available to warrant this. Additionally, questions such as whether such interventions are equally effective for different demographic groups remain unanswered.

Furthermore, a simple lack of benefit is not the same as a negative consequence and negative consequences should not be overlooked when assessing acceleration, particularly because it is such a hot-button issue within education. Many of the concerns around acceleration include not just the students who could be accelerated but also effects on the students who are not accelerated (e.g., loss of class leadership or role models), although one wonders if such questions are as much of a concern when interventions or grouping arrangements are employed to meet the needs of students with cognitive or learning disabilities (Peters & Matthews, 2016). Neihart (2007) concluded that grade-skipping, early entrance to college, and early school entrance have socio-affective benefits for students selected on the basis of academic readiness and social and emotional maturity, but also cautions that these programs may be harmful to individual students who are arbitrarily selected on the basis of IQ, suggesting that as with any educational strategy, implementation and the match between the intervention and the students' needs are keys to success.

Some research suggests that placing students in a more competitive, selective academic environment may result in a loss of self-concept—Big Fish Little Pond Effect (BFLPE; Marsh et al., 2015)—which can have a negative effect on future academic decisions. However, other research has shown that BFLPE may not influence high-performing students in the same way as other students (Makel, Lee, Olszewski-Kubliius, & Putallaz, 2012; Trautwein, Lüdtke, Marsh, & Nagy, 2009). Acceleration and grouping strategies likely differ in the extent to which BFLPE is evoked (e.g., more so when an individual student is accelerated vs. when students are cluster grouped within a classroom for advanced work). Investigations of the effects of acceleration cannot ignore the outcomes for nonaccelerated eligible students. However, providing evidence of a lack of difference continues to be difficult, especially when doing so does not fit within the traditional frequentist statistics framework (because one cannot affirm the null hypothesis) and requires Bayesian statistical comparisons that allow for confirmation of two groups performing equivalently.

Conclusion

Stanley (2000, p. 221) said that education should "avoid trying to teach students what they already know." Based on the nearly century's worth of research findings presented here, we believe that the data clearly suggest that ability grouping and acceleration are two such strategies for achieving this goal. The current findings will not settle all controversies on the philosophy of education. Nevertheless, we believe that they help clarify the academic effects of ability grouping and acceleration. Regardless, the conversation needs to evolve beyond whether such interventions can ever work. There is not an absence of evidence, nor is there evidence of absence of benefit. The preponderance of existing evidence accumulated over the past century suggests that academic acceleration and most forms of ability grouping like cross-grade subject grouping and special grouping for gifted students can greatly improve K–12 students' academic achievement.

References

References marked with an asterisk are ability grouping meta-analyses. References marked with a plus sign are acceleration meta-analyses included.

Adelson, J. L., McCoach, D. B., & Gavin, M. K. (2012). Examining the effects of gifted programming in mathematics and reading using the ECLS-K. *Gifted Child Quarterly*, *56*, 25–39. doi:10.1177/001698211431487

Ahn, S., Ames, A. J., & Myers, N. D. (2012). A review of meta-analyses in education: Methodological strengths and weaknesses. *Review of Educational Research*, *82*, 436–476. doi:10.3102/0034654312458162

Assouline, S. G., Colangelo, N., VanTassel-Baska, J., & Lupkowski-Shoplik, A. (2015). *A nation empowered: Evidence trumps the excuses holding back America's brightest students* (Vol. 2). Iowa City, IA: Belin-Blank Center.

Belfi, B., Goos, M., De Fraine, B., & Van Damme, J. (2012). The effect of class composition by gender and ability on secondary school students' school well-being and academic self-concept: A literature review. *Educational Research Review*, *7*, 62–74. doi:10.1016/j.edurev.2011.09.002

Bicak, L. J. (1962). *Achievement in eighth grade science by heterogeneous and homogeneous classes* (Unpublished doctoral dissertation). University of Minnesota, MN.

Bicak, L. J. (1964), Achievement in eighth grade science by heterogeneous and homogeneous classes. *Science Education*, *48*, 13–22. doi:10.1002/sce.3730480104

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

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Subgroup analyses. Introduction to meta-analysis*. Chichester, England: Wiley.

+Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2003). Comprehensive school reform and achievement. *Review of Educational Research*, *73*, 125–230. doi:10.3102/00346543073002125

Bui, S. A., Craig, S. G., & Imberman, S. A. (2011). *Is gifted education a bright idea? Assessing the impact of gifted and talented programs on achievement* (NBER Working Paper 17089). Retrieved from http://www.nber.org/papers/w17089

Callahan, C. M., Moon, T. R., & Oh, S. (2013). *National Survey of Gifted Programs*. Charlottesville: University of Virginia, National Research Center on the Gifted and Talented.

Chmielewski, A. K. (2014). An international comparison of achievement inequality in within- and between-school tracking systems. *American Journal of Education*, *120*, 293–324. doi:10.1086/675529

Chmielewski, A. K., Dumont, H., & Trautwein, U. (2013). Tracking effects depend on tracking type an international comparison of students' mathematics self-concept. *American Educational Research Journal*, *50*, 925–957 doi:10.3102/0002831213489843

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

Cooper, H., & Hedges, L. V. (1994). *The handbook of research synthesis and meta-analysis*. New York, NY: Russell Sage Foundation.

Cooper, H., & Koenka, A. C. (2012). The overview of reviews: Unique challenges and opportunities when research syntheses are the principal elements of new integrative scholarship. *American Psychologist*, *67*, 446–462. doi:10.1037/a0027119

Cox, C. M. (1926). The early mental traits of 300 geniuses. In *Genetic studies of genius* (Vol. 2). Stanford, CA: Stanford University Press.

Dewar, J. A. (1963). Grouping for arithmetic instruction in sixth grade. *The Elementary School Journal*, *63*(5), 266–269. Retrieved from http://www.jstor.org/stable/1000087

Drews, E. M. (1963). *Student abilities, grouping patterns, and classroom interaction* (Final Report of Cooperative Research Project No. 608, The Effectiveness of Homogeneous and Heterogeneous Ability Grouping in Ninth Grade English Classes with Slow, Average, and Superior Students). East Lansing: Michigan State University, Office of Research and Publications.

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*, 455–463. doi:10.1111/j.0006-341X.2000.00455.x/epdf

Fick, W. W. (1962). The effectiveness of ability grouping in seventh grade core classes. *Dissertation Abstracts International*, *23*, 2753. (University Microfilms No. 63-794)

Floyd, C. (1954). Meeting children's reading needs in the middle grades: A preliminary report. *Elementary School Journal*, *55*, 99−103. doi:10.1086/458653

Ford, S. (1974). *Grouping in mathematics: Effects on achievement and learning environment* (Unpublished doctoral dissertation). Yeshiva University, NY.

Gentry, P., Paul, K. A., McIntosh, J., Fugate, C. M., & Jen, E. (2014). *Total school cluster grouping and differentiation: A comprehensive research-based plan for raising student achievement and improving teacher practices*. Waco, TX: Prufrock Press.

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

*Goldring, E. B. (1990). Assessing the status of information on classroom organizational frameworks of gifted students. *Journal of Educational Research*, *83*, 313–326. doi:10.1080/00220671.1990.10885977

Gross, M. U. M. (1992). The use of radical acceleration in cases of extreme intellectual precocity. *Gifted Child Quarterly*, *36*, 90–98. doi:10.1177/001698629203600207

Gross, M. U. M. (2004). Radical acceleration. In S. G. Assouline, N. Colangelo, & J. VanTassel-Baska (Eds.), *A nation empowered* (Vol. 2, pp. 87–96). Iowa City, IA: Connie Belin and Jacqueline N. Blank International Center for Gifted Education and Talent Development.

Hattie, J. A. (2002). Classroom composition and peer effects. *International Journal of Educational Research*, *37*, 449–481. doi:10.1016/S0883-0355(03)00015-6

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational and Behavioral Statistics*, *6*, 107–128. doi:10.2307/1164588

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods*, *3*, 486–504. doi:10.1037/1082-989X.3.4.486

*Henderson, N. D. (1989). *A meta-analysis of ability grouping achievement and attitude in the elementary grades* (Unpublished doctoral dissertation). Mississippi State University, MS.

Higgins, J. P. T., & Green, S. (Eds.). (2011). *Cochrane handbook for systematic reviews of interventions* (Version 5.1.0). Retrieved from http://handbook.cochrane.org

Hillson, M., Jones, J. C, Moore, J. W., & Van Devender, F. (1964). A controlled experiment evaluating the effects of a nongraded organization on pupil achievement. *Journal of Educational Research*, *57*, 548−555.

Jones, C. J., Moore, J. W., & Van Devender, F. (1967). Comparison of pupil achievement after one and one-half and three years in a nongraded program. *Journal of Educational Research*, *61*, 75–77. doi:10.1080/00220671.1967.10883590

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*, 419-A. (University Microfilms No. AAT93-16362)

*Kulik, C. C. (1985, August). *Effects of inter-class ability grouping on achievement and self-esteem*. Paper presented at the 93rd annual convention of the American Psychological Association, Los Angeles, CA.

*Kulik, C. C., & Kulik, J. A. (1982). Effects of ability grouping on secondary school students: A meta-analysis of evaluation findings. *American Educational Research Journal*, *19*, 415–428. doi:10.3102/00028312019003415

*Kulik, C. C., & Kulik, J. A. (1984a, August). *Effects of ability grouping on elementary school pupils: A meta-analysis*. Paper presented at the annual meeting of the American Psychological Association, Toronto, Ontario, Canada.

+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* (Vol. 2, pp.13–22). Iowa City, IA: Connie Belin and Jacqueline N. Blank International Center for Gifted Education and Talent Development.

+Kulik, J. A., & Kulik, C. C. (1984b). Effects of accelerated instruction on students. *Review of Educational Research*, *54*, 409–425. doi:10.3102/00346543054003409

*Kulik, J. A., & Kulik, C. C. (1987). Effects of ability grouping on student achievement. *Equity & Excellence in Education*, *23*(1–2), 22–30. doi:10.1080/1066568870230105

*Kulik, J. A., & Kulik, C. C. (1992). Meta-analytic findings on grouping programs. *Gifted Child Quarterly*, *36*, 73–77. doi:10.1177/001698629203600204

*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*, 423–458. doi:10.3102/00346543066004423

Loveless, T. (2009). *Tracking and detracking: High achievers in Massachusetts middle schools*. Washington, DC: Thomas B. Fordham Institute.

Loveless, T. (2013). *How well are American students learning? The 2013 Brown Center report on American education*. Washington, DC: Brookings Institution.

Lubinski, D., & Benbow, C. P. (2000). States of excellence. *The American Psychologist*, *55*, 137–150. doi:10.1037/0003-066X.55.1.137

Makel, M. C., Lee, S. Y., Olszewski-Kubilius, P., & Putallaz, M. (2012). Changing the pond, not the fish: Following high ability students across different educational environments. *Journal of Educational Psychology*, *104*, 778–792. doi:10.1037/a0027558

Makel, M. C., Matthews, M. S., Peters, S. J., Rambo-Hernandez, K., & Plucker, J. A. (2016). *How can this so students be invisible? Large percentages of American students perform above grade level*. Retrieved from http://education.jhu.edu/edpolicy/commentary/IEPPolicyBriefFinalPLUCKER.pdf

Makel, M. C., & Plucker, J. A. (2014). Facts are more important than novelty: Replication in the education sciences. *Educational Researcher*, *43*, 304–316. doi:10.3102/0013189X14545513

Marascuilo, L. A., & McSweeney, M. (1972). Tracking and minority student attitudes and performance. *Urban Education*, *6*, 303–319. doi:10.1177/004208597200600401

Markow, D., & Cooper, M. (2008). *The MetLife Survey of the American Teacher: Past, present, and future*. Retrieved from http://files.eric.ed.gov/fulltext/ED504457.pdf

Marsh, H. W., Abduljabbar, A. S., Morin, A. J., Parker, P., Abdelfattah, F., Nagengast, B., & Abu-Hilal, M. M. (2015). The big-fish-little-pond effect: Generalizability of social comparison processes over two age cohorts from Western, Asian, and Middle Eastern Islamic countries. *Journal of Educational Psychology*, *107*, 258–271. doi:10.1037/a0037485

McClain, M. C., & Pfeiffer, S. (2012). Identification of gifted students in the United States today: A look at state definitions, policies, and practices. *Journal of Applied School Psychology*, *28*, 59–88. doi:10.1080/15377903.2012.643757

Mikkelson, J. E. (1962). *An experimental study of selective grouping and acceleration in junior high school mathematics* (Doctoral dissertation, University of Minnesota). *Dissertation Abstracts International*, *23*, 4226–4227.

Morgan, E. F., & Stucker, G. R. (1960). The Joplin Plan of reading vs. a traditional method. *Journal of Educational Psychology*, *51*, 69–73. doi:10.1037/h0047577

*Mosteller, F., Light, R. J., & Sachs, J. A. (1996). Sustained inquiry in education: Lessons from skill grouping and class size. *Harvard Educational Review*, *66*, 797–842. doi:10.17763/haer.66.4.36m328762x21610x

Neihart, M. (2007). The socioaffective impact of acceleration and ability grouping recommendations for best practice. *Gifted Child Quarterly*, *51*, 330–341. doi:10.1177/0016986207306319

*Noland, T. K. (1986). *The effects of ability grouping: A meta-analysis of research findings*. Retrieved from http://eric.ed.gov/?id=ED269451

Oakes, J. (1985). *Keeping track: How schools structure inequality*. New Haven, CT: Yale University Press.

Oakes, J. (2008). Keeping track: Structuring equality and inequality in an era of accountability. *Teachers College Record*, *110*, 700–712.

Olszewski-Kubilius, P. (2013, May 20). Setting the record straight on ability grouping. *Education Week*. Retrieved from http://www.edweek.org/tm/articles/2013/05/20/fp_olszewski.html

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2008). Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *Journal of Clinical Epidemiology*, *61*, 991–996. doi:10.1016/j.jclinepi.2007.11.010

Peters, S. J., & Matthews, M. S. (2016). Gifted education research from the economists' perspective: What have we learned? *Journal of Advanced Academics*, *27*, 150–161. doi:10.1177/1932202X16637398

Plucker, J. A., & Harris, B. (2015). Acceleration and economically vulnerable children. In S. G. Assouline, N. Colangelo, & J. VanTassel-Baska (Eds.), *A nation empowered* (Vol. 2, pp. 181–188). Iowa City, IA: Connie Belin and Jacqueline N. Blank International Center for Gifted Education and Talent Development.

Polanin, J. R., Maynard, B. R., & Dell, N. A. (2016). Overviews in educational research: A systematic review and analysis. *Review of Educational Research*. Advance online publication. doi:10.3102/0034654316631117

Pressey, S. L. (1946). Time-saving in professional training. *The American Psychologist*, *1*, 324–329. doi:10.1037/h0062575

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

Pressey, S. L. (1955). Concerning the nature and nurture of genius. *Scientific Monthly*, *81*, 123–129. Retrieved from http://www.jstor.org/stable/21966

+Rogers, 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. (2008). *Academic acceleration and giftedness: The research from 1990 to 2008. A best-evidence synthesis*. Paper presented at the proceedings of the Acceleration Poster Session at the 2008 Wallace Research Symposium on Talent Development. Retrieved from http://www.accelerationinstitute.org/proceedings_2008.pdf

Rogers, K. B. (2015). The academic, socialization, and psychological effects of acceleration: Research synthesis. In S. G. Assouline, N. Colangelo, & J. VanTassel-Baska (Eds.), *A nation empowered* (Vol. 2, pp. 19–29). Iowa City, IA: Connie Belin and Jacqueline N. Blank International Center for Gifted Education and Talent Development.

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). Publication bias in meta-analysis. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 1–9). Chichester, England: Wiley.

Ruhose, J., & Schwerdt, G. (2015). *Does early educational tracking increase migrant-native achievement gaps? Differences-in-differences evidence across countries* (CESifo Working Paper Series No. 5248). Retrieved from http://ssrn.com/abstract=2584887

Schmidt, F. L., & Oh, I.-S. (2013). Methods for second order meta-analysis and illustrative applications. *Organizational Behavior and Human Decision Processes*, *121*, 204–218. doi:10.1016/j.obhdp.2013.03.002

Seashore, C. E. (1922). The gifted student and research. *Science*, *56*, 641–648. doi:10.1126/science.56.1458.641

Shea, B. J., Hamel, C., Wells, G. A., Bouter, L. M., Kristjansson, E., Grimshaw, J., . . . Boers, M. (2009). AMSTAR is a reliable and valid measurement tool to assess the methodical quality of systematic reviews. *Journal of Clinical Epidemiology*, *62*, 1013–1020.

*Slavin, R. E. (1987). Ability grouping and student achievement in elementary schools: A best-evidence synthesis. *Review of Educational Research*, *57*, 293–336. doi:10.3102/00346543057003293

*Slavin, R. E. (1990). Achievement effects of ability grouping in secondary schools: A best-evidence synthesis. *Review of Educational Research*, *60*, 471–499. doi:10.3102/00346543060003471

*Slavin, R. E. (1993). Ability grouping in the middle grades: Achievement effects and alternatives. *Elementary School Journal*, *93*, 535–552. doi:10.1086/461739

Slavin, R. E., & Karweit, N. L. (1985). Effects of whole class, ability grouped, and individualized instruction on mathematics achievement. *American Educational Research Journal*, *22*, 351–367. doi:10.3102/00028312022003351

Smith, R. B. (1945). Pitfalls in acceleration. *Educational Research Bulletin*, *24*, 127–133. Retrieved from http://www.jstor.org/stable/1473019

Smith, W. M. (1961). The effect of intra-class ability grouping on arithmetic achievement in grades two through five. *Dissertation Abstracts International*, *21*, 563–556.

Southern, W. T., & Jones, E. D. (2015). Types of acceleration: Dimensions and issues. In S. G. Assouline, N. Colangelo, J. VanTassel-Baska, & A. Lupkowski-Shoplik (Eds.), *A nation empowered* (Vol. 2, pp. 9–18). Iowa City, IA: Connie Belin and Jacqueline N. Blank International Center for Gifted Education and Talent Development.

Southern, W. T., Jones, E. D., & Fiscus, E. D. (1989). Practitioner objections to the academic acceleration of gifted children. *Gifted Child Quarterly*, *33*, 29–35. doi:10.1177/001698628903300105

Stanley, J. C. (2000). Helping students learn only what they don't already know. *Psychology, Public Policy, and Law*, *6*, 216–222. doi:10.1037/1076-8971.6.1.216

Stanley, T. D. (2005). Beyond publication bias. *Journal of Economical Surveys*, *19*, 309–345. doi:10.1111/j.0950-0804.2005.00250.x

Stanley, T. D., & Doucouliagos, H. (2007). *Identifying and correcting publication selection bias in the efficiency-wage literature: Heckman meta-regression*. Retrieved from https://ideas.repec.org/p/dkn/econwp/eco_2007_11.html

Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods*, *5*, 60–78. doi:10.1002/jrsm.1095

+Steenbergen-Hu, S., & Moon, S. (2011).The effects of acceleration on high-ability learners: A meta-analysis. *Gifted Child Quarterly*, *55*, 39–53. doi:10.1177/0016986210383155

Terman, L. (1925). *Genetic studies of genius: Vol. 1. The mental and physical traits of a thousand gifted children*. Stanford, CA: Stanford University Press.

Tieso, C. L. (2003). Ability grouping is not just tracking anymore. *Roeper Review*, *26*, 29–36. doi:10.1080/02783190309554236

Trautwein, U., Lüdtke, O., Marsh, H. W., & Nagy, G. (2009). Within school social comparison: How students perceive the standing of their class predicts academic self-concept. *Journal of Educational Psychology*, *101*, 853–866. doi:10.1037/a0016306

U.S. Department of Education. (2014). *Public and private elementary and secondary teachers, enrollment, pupil/teacher ratios, and new teacher hires: Selected years, fall 1955 through fall 2023* (Table 208.20). Retrieved from https://nces.ed.gov/programs/digest/d13/tables/dt13_208.20.asp

Wallen, N. E., & Vowles, R. O. (1960). The effect of intraclass ability grouping on arithmetic achievement in the sixth grade. *Journal of Educational Psychology*, *51*, 159–163. doi:10.1037/h0040117

Whitehurst, G. (2003). *The Institute of Education Sciences: New wine, new bottles. A presentation by IES director Grover (Russ) Whitehurst*. Retrieved from http://ies.ed.gov/director/speeches2003/04_22/2003_04_22.asp

Authors

SAIYING STEENBERGEN-HU, PhD, is a research assistant professor and the research director of the Center for Talent Development of Northwestern University's School of Education and Social Policy, 617 Dartmouth Place, Evanston, IL 60208; email: *hu@northwestern.edu*. Her research focuses on the effectiveness of educational interventions on students' academic achievement and psychosocial development. Her methodological interests and skills include meta-analysis, research synthesis, educational measurement and assessment, quantitative research methodology, and applied statistical analysis. Her meta-analysis, coauthored with Sidney Moon, on the effects of acceleration on high-ability learners won the *Gifted Child Quarterly* Paper of the Year Award in 2012. She and Paula Olszewski-Kubilius recently published a methodological brief, "How to Conduct a Good Meta-Analysis in Gifted Education," in *Gifted Child Quarterly* (Volume 60, No. 2, 2016).

MATTHEW C. MAKEL, PhD, is the director of research at the Duke University Talent Identification Program, 300 Fuller St., Durham, NC 27701; email: *mmakel@tip.duke.edu*. His research focuses on research methods and the nature and development of the abilities, perceptions, and environments of academically talented youth.

PAULA OLSZEWSKI-KUBILIUS, PhD, is the director of the Center for Talent Development at Northwestern University and a professor in the School of Education and Social Policy, 617 Dartmouth Place, Evanston, IL 60208; email: *p-olszewski-kubilius@northwestern.edu*. Over the past 30 years, she has created programs for all kinds of gifted learners and written extensively about talent development. She has served as the editor of *Gifted Child Quarterly*, coeditor of the *Journal of Secondary Gifted Education*, and on the editorial boards of *Gifted and Talented International, Roeper Review*, and *Gifted Child Today*. She is past-president of the National Association for Gifted Children and received the Distinguished Scholar Award in 2009 from the National Association for Gifted Children.