Exhibit 1.8

(page intentionally left blank)

Review of Educational Research
Spring 2006, Vol. 76, No. 1, pp. 1–62

Does Homework Improve Academic Achievement? A Synthesis of Research, 1987–2003

Harris Cooper, Jorgianne Civey Robinson, and Erika A. Patall
Duke University

In this article, research conducted in the United States since 1987 on the effects of homework is summarized. Studies are grouped into four research designs. The authors found that all studies, regardless of type, had design flaws. However, both within and across design types, there was generally consistent evidence for a positive influence of homework on achievement. Studies that reported simple homework–achievement correlations revealed evidence that a stronger correlation existed (a) in Grades 7–12 than in K–6 and (b) when students rather than parents reported time on homework. No strong evidence was found for an association between the homework–achievement link and the outcome measure (grades as opposed to standardized tests) or the subject matter (reading as opposed to math). On the basis of these results and others, the authors suggest future research.

Keywords: homework, meta-analysis.

Homework can be defined as any task assigned by schoolteachers intended for students to carry out during nonschool hours (Cooper, 1989). This definition explicitly excludes (a) in-school guided study; (b) home study courses delivered through the mail, television, audio or videocassette, or the Internet; and (c) extracurricular activities such as sports and participation in clubs. The phrase "intended for students to carry out during nonschool hours" is used because students may complete homework assignments during study hall, library time, or even during subsequent classes.

Variations in homework can be classified according to its (a) amount, (b) skill area, (c) purpose, (d) degree of choice for the student, (e) completion deadline, (f) degree of individualization, and (g) social context. Variations in the *amount* of homework can appear as differences in both the frequency and length of individual assignments. Assignments can range over all the *skill areas* taught in school.

The *purposes* of homework assignments can be divided into (a) instructional and (b) noninstructional objectives (cf. Epstein, 1988, 2001; Epstein & Van Voorhis, 2001; Lee & Pruitt, 1979). The most common instructional purpose of homework is to provide the student with an opportunity to practice or review material that has already been presented in class (Becker & Epstein, 1982). Preparation assignments introduce material to help students obtain the maximum benefit when the new material is covered in class (Muhlenbruck, Cooper, Nye, & Lindsay, 1999). Extension homework involves the transfer of previously learned skills to new situations

(Lee & Pruitt, 1979). Finally, homework can require students to integrate separately learned skills and concepts (Lee & Pruitt, 1979). This might be accomplished using book reports, science projects, or creative writing.

Homework has other purposes in addition to enhancing instruction. It can be used to (a) establish communication between parent and child (Acock & Demo, 1994; Balli, Demo, & Wedman, 1998; Epstein, Simon, & Salinas, 1997; González, Andrade, Civil, & Moll, 2001; Scott-Jones, 1995; Van Voorhis, 2003); (b) fulfill directives from school administrators (Hoover-Dempsey, Bassler, & Burow, 1995); and (c) punish students (Epstein & Van Voorhis, 2001; Xu & Corno, 1998). To this list might be added the public relations objective of simply informing parents about what is going on in school (Coleman, Hoffer, & Kilgore, 1982; Corno, 1996; Rutter, Maughan, Mortimore, & Ouston, 1979).

Homework assignments rarely reflect a single purpose. Rather, most assignments serve several different purposes; some relate to instruction, whereas others may meet the purposes of the teacher, the school administration, or the school district.

The *degree of choice* afforded a student refers to whether the homework assignment is compulsory or voluntary. Related to the degree of choice, *completion deadlines* can vary from short term, meant to be completed overnight or for the next class meeting, to long term, with students given days or weeks to complete the task. The *degree of individualization* refers to whether the teacher tailors assignments to meet the needs of each student or whether a single assignment is presented to groups of students or to the class as a whole. Finally, homework assignments can vary according to the *social context* in which they are carried out. Some assignments are meant for the student to complete independent of other people. Assisted homework explicitly calls for the involvement of another person, a parent or perhaps a sibling or friend. Still other assignments involve groups of students working cooperatively to produce a single product.

Overview

The Importance of Homework and Homework Research

Homework is an important part of most school-aged children's daily routine. According to the National Assessment of Educational Progress (Campbell et al., 1996), over two-thirds of all 9-year-olds and three-quarters of all 13- and 17-year-olds reported doing some homework every day. Sixteen percent of 9-year-olds reported doing more than 1 hour of homework each day, and this figure jumped to 37% for 13-year-olds and 39% for 17-year-olds. More recent surveys support the extensive use of homework, although the amount of homework that students report varies from study to study, depending perhaps on how the question is asked. For example, Gill and Schlossman (2003) reported recent declines in time spent on homework. However, among the youngest students, age 6 to 8, homework appears to have increased between 1981 (52 minutes weekly) and 1997 (128 minutes weekly; Hofferth & Sandberg, 2000).

Homework likely has a significant impact on students' educational trajectories. Most educators believe that homework can be an important supplement to in-school academic activities (Henderson, 1996). However, it is also clear from the surveys mentioned earlier that not all teachers assign homework and/or not all students complete the homework they are assigned. This suggests that whatever impact homework might have on achievement varies from student to student, depending on how much each student is assigned or completes.

Homework is often a source of friction between home and school. Accounts of conflicts between parents and educators appear often in the popular press (e.g., Ratnesar, 1999; Coutts, 2004; Kralovec & Buell, 2000; Loveless, 2003). Parents protest that assignments are too long or too short, too hard or too easy, or too ambiguous (Baumgartner, Bryan, Donahue, & Nelson, 1993; Kralovec & Buell, 2000; Warton, 1998). Teachers complain about a lack of support from parents, a lack of training in how to construct good assignments, and a lack of time to prepare effective assignments (Farkas, Johnson, & Duffet, 1999). Students protest about the time that homework takes away from leisure activities (Coutts, 2004; Kralovec & Buell, 2000). Many students consider homework the chief source of stress in their lives (Kouzma & Kennedy, 2002).

To date, the role of research in forming homework policies and practices has been minimal. This is because the influences on homework are complex, and no simple, general finding applicable to all students is possible. In addition, research is plentiful enough that a few studies can always be found to buttress whatever position is desired, while the counter-evidence is ignored. Thus advocates for or against homework often cite isolated studies either to support or to refute its value.

It is critical that homework policies and practices have as their foundation the best evidence available. Policies and practices that are consistent with a trustworthy synthesis of research will (a) help students to obtain the optimum education benefit from homework, and (b) help parents to find ways to integrate homework into a healthy and well-rounded family life. It is our intention in this article to collect as much of the research as possible on the effects of homework, both positive and negative, conducted since 1987. We will apply narrative and quantitative techniques to integrate the results of studies (see Cooper, 1998; Cooper & Hedges, 1994). While research rarely, if ever, covers the gamut of issues and circumstances confronted by educators, we hope that the results of this research synthesis will be used both to guide future research on homework and to assist in the development of policies and practices consistent with the empirical evidence.

A Brief History of Homework in the United States

Public attitudes toward homework have been cyclical (Gill & Schlossman, 1996, 2004). Prior to the 20th century, homework was believed to be an important means for disciplining children's minds (Reese, 1995). By the 1940s, a reaction against homework had set in (Nash, 1930; Otto, 1941). Developing problem-solving abilities, as opposed to learning through drill, became a central task of education (Lindsay, 1928; Thayer, 1928). Also, the life-adjustment movement viewed home study as an intrusion on other at-home activities (Patri, 1925; San Diego City Schools Research Department, 1936).

The trend toward less homework was reversed in the late 1950s after the Russians launched the Sputnik satellite (Gill & Schlossman, 2000; Goldstein, 1960; Epps, 1966). Americans became concerned that a lack of rigor in the educational system was leaving children unprepared to face a complex technological future and to compete against our ideological adversaries. Homework was viewed as a means of accelerating the pace of knowledge acquisition. But in the mid-1960s the cycle again reversed itself (Jones & Colvin, 1964). Homework came to be seen as a symptom of excessive pressure on students. Contemporary learning theories again questioned the value of homework and raised its possible detrimental consequences for mental health.

By the mid-1980s, views of homework had again shifted toward a more positive assessment (National Commission on Excellence in Education, 1983). In the wake of declining achievement test scores and increased concern about American's ability to compete in a global marketplace, homework underwent its third renaissance in 50 years. However, as the century turned, and against the backdrop of continued parental support for homework (Public Agenda, 2000), a predicable backlash set in, led by beleaguered parents concerned about the stresses on their children (Winerip, 1999).

Past Syntheses of Homework Research

Homework has been an active area of study among American education researchers for the past 70 years. As early as 1927, a study by Hagan (1927) compared the effects of homework with the effects of in-school supervised study on the achievement of 11- and 12-year-olds. However, researchers have been far from unanimous in their assessments of the strengths and weaknesses of homework. For example, more than a dozen reviews of the homework literature were conducted between 1960 and 1987 (see Cooper, 1989, for a detailed description). The conclusions of these reviews varied greatly, partly because they covered different literature, used different criteria for inclusion of studies, and applied different methods for the synthesis of study results.

Cooper (1989) conducted a review of nearly 120 empirical studies of homework's effects and the ingredients of successful homework assignments. Quantitative synthesis techniques were used to summarize the literature. This review included three types of studies that help answer the general question of whether homework improves students' achievement. The first type of study compared achievement of students given homework assignments with students given no homework. In 20 studies conducted between 1962 and 1986, 14 produced effects favoring homework while 6 favored no homework. Most interesting was the influence of grade level on homework's relation with achievement. These studies revealed that the average high school student in a class doing homework outperformed 69% of the students in a no-homework class, as measured by standardized tests or grades. In junior high school, the average homework effect was half this magnitude. In elementary school, homework had no association with achievement gains.

The next type of evidence compared homework with in-class supervised study. Overall, the positive effect of homework was about half what it was when students doing homework were compared with those not doing homework. Most important was the emergence once again of a strong grade-level effect. When homework and in-class study were compared in elementary schools, in-class study proved superior.

Finally, Cooper found 50 studies that correlated the amount of time students spent on homework with a measure of achievement. Many of these correlations came from statewide surveys or national assessments. In all, 43 correlations indicated that students who did more homework had better achievement outcomes, while only 7 indicated negative outcomes. Again, a strong grade-level interaction appeared. For students in elementary school, the average correlation between amount of homework and achievement was nearly $r = 0$; for students in middle grades it was $r = .07$; and for high school students it was $r = .25$.

The Need for a New Synthesis of the Homework Literature

There are three reasons for conducting a new synthesis of the homework literature: (a) to update the evidence on past conclusions about the effects of homework and determine if the conclusions from research need modification; (b) to determine whether some of the questions left unanswered by the earlier syntheses can now be answered; and (c) to apply new research synthesis techniques.

In the years since the completion of Cooper's (1989) meta-analysis, a substantial new body of evidence has been added to the homework literature. For example, a search of ERIC, PsycINFO, Sociological Abstracts, and Dissertation Abstracts between January 1987 (when the search for the earlier synthesis ended) and December 2003 indicated that over 4,000 documents with homework as a keyword had been added to these reference databases. When we delimited this search to documents that the reference engine cataloged as "empirical," nearly 900 documents remained. Yet we know of no comprehensive attempt to synthesize this new literature. Therefore, a reassessment of the evidence seems timely, both to determine if the earlier conclusions need to be modified and to benefit from the added precision that the new evidence can bring to the current assessment.

Cooper's meta-analysis revealed a consistent influence of grade level on the homework–achievement relationship. However, it produced ambiguous results regarding the possible differential impact of homework on different subject matters and on different measures of achievement. Specifically, research using different comparison groups (i.e., no homework, supervised study, correlations involving different reported amounts of homework) produced different orderings or magnitudes of homework's relation to achievement for different subject matters and achievement measures. Also, Cooper (1989) found uniformly nonsignificant relationships between the sex of the student and the magnitude of the homework–achievement relationship. However, some recent theoretical perspectives (Covington, 1998; Deslandes & Cloutier, 2002; Harris, Nixon, & Rudduck, 1993; Jackson, 2003) suggest that girls generally hold more positive attitudes than boys toward homework and expend greater effort on it. Emerging evidence from some homework studies (Harris et al., 1993; Hong & Milgram, 1999; Younger & Warrington, 1996) lends empirical support to these perspectives.

While these theories and results do not directly predict a stronger *relationship* between homework and achievement for girls than for boys (that is, they predict a main effect of higher levels of achievement among girls than among boys but do not indicate why differences in homework attitude and effort within the sexes would be more closely tied to achievement for one sex than the other, an interaction effect), they do suggest that this remains an important issue. Therefore, exploring these moderating relationships will be a focus of the present synthesis.

Also, the Cooper (1989) synthesis paid only passing attention to the ability of the cumulated evidence to establish a causal relationship between homework and achievement. Clearly, the 50 studies that took naturalistic, cross-sectional measures of the amount of time students spent on homework and correlated these with measures of achievement cannot be used to establish causality. About half of the studies that introduced homework as an exogenous intervention and then compared achievement for students who did homework with that of students who did not, or who had in-school supervised study, employed random assignment of students to conditions. The other half sometimes did and sometimes did not employ a priori matching or post hoc statistical equating to enhance the similarity of homework and no-homework groups. When homework was compared with no-homework, Cooper reported that studies that used random assignment produced positive effects of homework similar to nonrandom assignment studies. However, when compared with in-school supervised study, random-assignment designs revealed no difference between the homework and in-school study students. We will test to determine whether these findings still hold for the new evidence.

Also, since the earlier synthesis appeared, numerous studies have employed structural equation modeling to test the fit of complex models of the relationship between various factors and student achievement. Homework has been used as a factor in many of these models. The earlier synthesis did not include these designs, but this synthesis will.

Methodologically, the past two decades have introduced new techniques and refinements in the practice of research synthesis. These include, among others, two important advances. First, there is now a greater understanding of meta-analytic error models involving the use of fixed and random-error assumptions that add precision to statements about the generality of findings. Second, new tests have been developed to estimate the impact of data censoring on research synthesis findings. These give us a better sense of the robustness of findings against plausible missing data assumptions. We will use these in the synthesis that follows.

Potential Measures of the Effects of Homework

As might be expected, educators have suggested a long list of both positive and negative consequences of homework (Cooper, 1989; see also Epstein, 1988; Warton, 2001). Table 1 presents a list of potential outcomes that might be the focus of homework research and the potential measures of interest for this synthesis.

The positive effects of homework can be grouped into four categories: (a) immediate achievement and learning; (b) long-term academic; (c) nonacademic; and, (d) parental and family benefits. The immediate effect of homework on learning is its most frequent rationale. Proponents of homework argue that it increases the time students spend on academic tasks (Carroll, 1963; Paschal, Weinstein, & Walberg, 1984; Walberg & Paschal, 1995). Thus the benefits of increased instructional time should accrue to students engaged in home study. The long-term academic benefits of homework are not necessarily enhancements to achievement in particular academic domains, but rather the establishment of general practices that facilitate learning. Homework is expected to (a) encourage students to learn during their leisure time; (b) improve students' attitudes toward school; and (c) improve students' study habits and skills (Alleman & Brophy, 1991; Corno & Xu, 1998; Johnson & Pontius, 1989; Warton, 2001).

Also, homework has been offered as a means for developing personal attributes in children that can promote positive behaviors that, in addition to being important for academic pursuits, generalize to other life domains. Because homework generally requires students to complete tasks with less supervision and under less severe time constraints than is the case in school, home study is said to promote greater self-

TABLE 1
Potential effects of homework that might serve as outcomes for research

Potential positive effects
Immediate achievement and learning
Better retention of factual knowledge
Increased understanding
Better critical thinking, concept formation, information processing
Curriculum enrichment
Long-term academic benefits
More learning during leisure time
Improved attitude toward school
Better study habits and skills
Nonacademic benefits
Greater self-direction
Greater self-discipline
Better time organization
More inquisitiveness
More independent problem-solving
Parental and family benefits
Greater parental appreciation of and involvement in schooling
Parental demonstrations of interest in child's academic progress
Student awareness of connection between home and school
Potential negative effects
Satiation
Loss of interest in academic material
Physical and emotional fatigue
Denial of access to leisure time and community activities
Parental interference
Pressure to complete homework and perform well
Confusion of instructional techniques
Cheating
Copying from other students
Help beyond tutoring
Increased differences between high and low achievers

Note. Adapted from Cooper (1989).

direction and self-discipline (Corno, 1994; Zimmerman, Bonner, & Kovach, 1996), better time organization, more inquisitiveness, and more independent problem solving. These skills and attributes apply to the nonacademic spheres of life as well as the academic.

Finally, homework may have positive effects on parents and families (Hoover-Dempsey et al., 2001). Teachers can use homework to increase parents' appreciation of and involvement in schooling (Balli, 1998; Balli, Wedman, & Demo, 1997; Epstein & Dauber, 1991; Van Voorhis, 2003). Parents can demonstrate an interest in the academic progress of their children (Epstein & Van Voorhis, 2001; Balli, Demo,

& Wedman, 1998). Students become aware of the connection between home and school.

Some negative effects attributed to homework contradict the suggested positive effects. For instance, opponents of homework have argued that it can have a negative influence on attitudes toward school (Chen, & Stevenson, 1989), by satiating students on academic pursuits. They claim any activity remains rewarding for only so long, and children may become overexposed to academic tasks (Bryan, Nelson, & Mathru, 1995). Related to the satiation argument is the notion that homework leads to general physical and emotional fatigue. Homework can also deny children access to leisure time and community activities (Warton, 2001; Coutts, 2004). Proponents of leisure activities point out that homework is not the only circumstance under which after-school learning takes place. Many leisure activities teach important academic and life skills.

Involving parents in the schooling process can have negative consequences (Epstein, 1988; Levin, Levy-Shiff, Appelbaum-Peled, Katz, Komar, & Meiran, 1997; Cooper, Lindsay, & Nye, 2000). Parents pressure students to complete homework assignments or to do them with unrealistic rigor. Also, parents may create confusion if they are unfamiliar with the material that is sent home for study or if their approach to teaching differs from that used in school. Parental involvement—indeed the involvement of anyone else in homework—can sometimes go beyond simple tutoring or assistance. This raises the possibility that homework might promote cheating or excessive reliance on others for help with assignments.

Finally, some opponents of homework have argued that home study has increased differences between high- and low-achieving students, especially when the achievement difference is associated with economic differences (Scott-Jones, 1984; Odum, 1994; McDermott, Goldman, & Varenne, 1984). They suggest that high achievers from well-to-do homes will have greater parental support for home study, including more appropriate parental assistance. Also, these students are more likely to have access to places conducive to their learning style in which to do assignments and better resources to help them complete assignments successfully.

With few exceptions, the positive and negative consequences of homework can occur together. For instance, homework can improve study habits at the same time that it denies access to leisure-time activities. Some types of assignments can produce positive effects, whereas other assignments produce negative ones. In fact, in light of the host of ways that homework assignments can be construed and carried out, complex patterns of effects ought to be expected.

The present synthesis will search for any and all of the above possible effects of homework. However, it is unrealistic to expect that any but a few of these will actually appear in the research literature. We expected the large preponderance of measures to involve achievement test scores, school grades, and unit grades. A few measures of students' attitudes toward school and subject matters might also appear. Other measures of homework's effect were expected to be few and far between. One reason for this is because many of the other potential effects are subtle. Therefore, their impact might take a long time to accrue, and few researchers have the resources to mount and sustain long-term longitudinal research. Another reason for the lack of subtle measures of homework's effect is that the homework variable is often one of many influences on achievement being examined in a study. It is achievement as the outcome that is the primary focus of investigation with many predictors, rather than homework as the focus with many outcomes measured.

Factors That Affect the Utility of Homework Assignments

In addition to looking at homework's effectiveness on different outcomes, researchers have examined how other variations in assignments might influence their utility. Homework assignments are influenced by more factors than any other instructional strategy. Student differences may play a major role because homework allows students considerable discretion about whether, when, and how to complete assignments. Teachers may structure and monitor homework in a multitude of ways. The home environment may influence the process by creating a positive or negative atmosphere for study. And finally, the broader community provides other leisure activities that compete for the student's time.

Table 2 presents a model of the homework process presented by Cooper (1989). The model organizes into a single scheme many of the factors that educators have suggested might influence the success of a homework assignment. The model proposes that student ability, motivation, and grade level, as well as other individual differences (e.g., sex, economic background), and the subject matter of the homework assignments are exogenous factors, or moderator conditions, that might influence homework's effect. The model's endogenous factors, or mediators, divide the homework process into characteristics of the assignment and a home-community phase sandwiched by two classroom phases, each containing additional potential influences on homework's effects. Finally, Table 2 includes the potential consequences of homework as the outcomes in the process.

In this synthesis, the search for factors that might influence the impact of homework will focus only on the exogenous factors and the outcome variables, with the exception of the endogenous factor of amount of homework. Studies of the latter type are included because (a) they would include students who did no homework at all; and (b) achievement variations related to time spent on homework can reasonably be taken to bear on homework's effectiveness. Our restriction is based on the fact that most studies that look at other variations in endogenous or mediating factors rarely do so in the context of an investigation that also attempts to assess the more general effects of homework. Investigations of mediating factors typically pit one homework strategy against another and do not contain a condition in which students receive no-homework or an alternative treatment. Thus, in an effort to keep our task manageable, we focused here on studies that investigate primarily the general effects of homework, and we excluded studies that exclusively examine variations in homework assignments. (For a review of one such endogenous variable, parent involvement, see Patall, Cooper, & Robinson, 2005).

Optimum Amounts of Homework

Related to the issue of time spent on homework is the important question concerning the optimum amount of homework. Cooper (1989) found nine studies that allowed for a charting of academic performance as a function of homework time. The line-of-progress was flat in young children. For junior high school students, achievement continued to improve with more homework until assignments lasted between 1 and 2 hours a night. More homework than that was no longer associated with higher achievement. For high school students, the line-of-progress continued to

TABLE 2

A temporal model of factors influencing the effects of homework

Exogenous factors	Assignment characteristics	Initial classroom factors	Home–community factors	Classroom follow-up	Outcomes or effects
Student characteristics	Amount	Provision of materials	Competitors for student time	Feedback	Assignment completion
Ability	Purpose	Facilitators	Home environment	Written comments	Assignment performance
Motivation	Skill area used	Suggested approaches	Space	Grading	Positive effects
Study habits	Degree of individualization	Links to curriculum	Light	Incentives	Immediate academic
Subject matter	Degree of student choice	Other rationales	Quiet	Testing of related content	Long-term academic
Grade level	Completion deadlines		Materials	Use in class discussion	Nonacademic
	Social context		Others' involvement		Parental and family
			Parents		Negative effects
			Siblings		Satiation
			Other students		Limited access to leisure and community activities
					Parental interference
					Cheating
					Increased student differences

Note. From Cooper (1989). Copyright 2005 by American Psychological Association. Reprinted with permission.

go up through the highest point on the measured scales, more than 2 hours. In the present synthesis, we included studies examining time on homework because of their relevance to homework's general effectiveness; therefore, we also looked for studies that might replicate or extend this finding.

Bias and Generalization in Research Synthesis

Decisions concerning how to search the literature determine the kinds of materials that will form the basis for a synthesis' conclusions. Identifying the literature is complicated by the fact that the search has two targets (Cooper, 1998). First, synthesists want to locate all previous literature on the problem. This is especially critical with regard to the retrieval of studies for inclusion in a meta-analysis. Synthesists can exert some control over whether this goal is achieved through their choice of information sources. Second, synthesists hope that the included studies will allow generalizations in the broader topic area. The generalizability of our synthesis was constrained by the students, schools, and communities represented in the literature.

We employed several strategies to ensure that our homework synthesis included the most exhaustive set of relevant documents. These strategies included (a) computerized searches of reference databases; (b) direct contact with active researchers and others who might know of unpublished or "fugitive" homework research; and (c) scrutiny of reference lists of relevant materials. In addition, analyses of the retrieved studies were undertaken to test for indications that the studies in hand might constitute a biased representation of the population of studies, and if so, to determine the nature of the bias.

Avoiding overgeneralization requires recognizing that the students, schools, and communities represented in the retrieved literature may not represent all target populations. For instance, it may be that little or no research has been conducted that examines the effects of homework on first- or second-grade students. A synthesis that qualifies conclusions with information about the kinds of people missing or overrepresented in studies runs less risk of overgeneralization. Such an examination of potential population restrictions will be included in the present work.

Methods for Research Synthesis

Literature Search Procedures

No matter how thorough the procedures may be, no search of the literature is likely to succeed in retrieving all studies relating homework to achievement. Therefore, systematic data censoring is a concern. That is, the possibility exists that more easily retrievable studies have different results from studies that could not be retrieved. To address this possibility, we collected studies from a wide variety of sources and included search strategies meant to uncover both published and unpublished research.

First, we searched the ERIC, PsycINFO, Sociological Abstracts, and Dissertation Abstracts electronic databases for documents cataloged between January 1, 1987, and December 31, 2003. The single keyword "homework" was used in these searches. Also, the Science Citation Index Expanded and the Social Sciences Citation Index databases were searched from 1987 to 2004 to identify studies or reviews that had cited Cooper (1989). These searches identified approximately 4,400 nonduplicate potentially relevant studies.

Next, we employed three direct-contact strategies to ensure that we tapped sources that might have access to homework-related research that would not be included in the reference and citation databases. First, we contacted the dean, associate dean, or chair of 77 colleges, schools, or departments of education at research-intensive institutions of higher education and requested that they ask their faculty to share with us any research they had conducted that related to the practice of assigning homework. Second, we sent similar letters to 21 researchers who, as revealed by our reference database search, had been the first author on two or more articles on homework and academic achievement between 1987 and the end of 2003. Finally, we sent similar letters to the directors of research or evaluation in more than a hundred school districts, obtained from the membership list of the National Association of Test Directors.

Two researchers in our team then examined each title, abstract, or document. If either of the two felt that the document might contain data relevant to the relationship between homework and an achievement-related outcome, we obtained the full document (in the case of judgments made on the titles or abstracts).

Finally, the reference sections of relevant documents were examined to determine if any cited works had titles that also might be relevant to the topic.

Criteria for Including Studies

For a study to be included in the research synthesis, several criteria had to be met. Most obviously, the study had to have estimated in some way the relationship between a measure of homework activity on the part of a student and a measure of achievement or an achievement-related outcome.

Two sampling restrictions were placed on included studies. Each study had to assess students in kindergarten through 12th grade. We excluded studies conducted on preschool-aged children or on postsecondary students. It was felt that the purpose and causal structure underlying the homework–achievement relationship would be very different for these populations. For similar reasons, we included only studies conducted in the United States.

Finally, the report had to contain enough information to permit the calculation of an estimate of the homework–achievement relationship.

Information Retrieved From Evaluations

Numerous characteristics of each study were included in the database. These characteristics encompassed six broad distinctions among studies: (a) the research report; (b) the research design; (c) the homework variable; (d) the sample of students; (e) the measure of achievement, and (f) the estimate of the relationship between homework and achievement.

Report Characteristics

Each database entry began with the name of the author of the study. Then the year of the study was recorded, followed by the type of research report. Each research report was categorized as a journal article, book chapter, book, dissertation, Master's thesis, private report, government report (state or federal), school or district report, or other type of report.

Research Design and Other Study Characteristics

The studies in this research synthesis were categorized into three basic design types, some with subtypes.

First, studies could employ exogenous manipulations of homework. This meant that the presence or absence of homework assignments was manipulated expressly for purposes of the study. Within the exogenous manipulation studies, the experimenters could introduce the manipulation at the student or classrooms level, either by randomly assigning students to homework and no-homework conditions or by some nonrandom process. If a nonrandom process was used, the experimenter then might or might not employ a priori matching or post hoc statistical procedures to equate the homework and no-homework groups. If procedures were used to equate groups, the variables used to enhance the equivalence of the groups could differ from study to study. Each of these variations in design was recorded for the set of studies that used exogenous homework manipulations.

In addition to these design characteristics of exogenous homework manipulation studies and their report information, we recorded (a) the number of students and classrooms included in the homework and no-homework conditions at the beginning and end of the experiment; (b) the grade level of the students; (c) the subject matter of the homework (reading, other language arts, math, science, social studies, foreign language, other, or multiple subjects); (d) the number of assignments per week and their duration; (e) the measure of achievement (standardized achievement test, teacher-developed unit test, textbook chapter unit test, class grades, overall grade point average, composite achievement score); and (f) the magnitude of the relationship between homework and achievement.

The second type of design included studies that took naturalistic, cross-sectional measures of the amount of time the students spent on homework without any intervention on the part of the researchers and related these to an achievement-related measure. This second type of design also included an attempt to statistically equate students on other variables that might be confounded with homework and therefore might account for the homework–achievement relationship. For these studies, we also coded the source of the data, that is, whether the data were collected by the researchers or by an independent third party. If data were from an independent source, we coded the source. We coded the analytic strategy used to equate students. Most frequently, this involved conducting multiple regression analysis or the application of a structural equation modeling package. Also, we coded each of the same variables coded for studies that used exogenous manipulations of homework, except for (a) the sample sizes in the homework and no-homework groups (only total sample size in the analysis was recorded); and (b) the number and duration of assignments, which was irrelevant to this design. Instead of the assignment characteristics, we coded the amount of time the student spent doing homework, as measured by student or parent report.

The third type of design involved the calculation of a simple bivariate correlation between the time the student spent on homework and the measure of achievement. In these studies, no attempt was made to equate students on other variables that might be confounded with time on homework. For these studies, we also recorded the same variables coded for studies using statistical controls of other variables except, of course, the number and nature of controlled variables. We also coded several additional variables related to the sample of students. These included the students' (a) sex; (b) socioeconomic status (low, low-middle, middle, middle-upper, upper, "mixed," no SES [socioeconomic status] information given); and (c) whether any of the following labels were applied to the sample of students (gifted, average, "at risk," underachieving/below grade level, possessing a learning disability, overachieving/above grade level).

Effect Size Estimation

For studies with exogenous manipulations of homework, we used the standardized mean difference to estimate the effect of homework on measures of student achievement. The *d*-index (Cohen, 1988) is a scale-free measure of the separation between two group means. Calculating the *d*-index for any comparison involves dividing the difference between the two group means by either their average standard deviation or by the standard deviation of the control group. This calculation results in a measure of the difference between the two group means expressed in terms of their common standard deviation or that of the untreated population. Thus a *d*-index of .25 indicates that one-quarter standard deviation separates the two means. In the synthesis, we subtracted the no-homework condition mean from the homework condition mean and divided the difference by their average standard deviation. Thus positive effect sizes indicate that the students doing homework had better achievement outcomes.

We calculated effect sizes based on the means and standard deviations of students' achievement indicators, if available. If means and standard deviations were not available, we retrieved the information needed from inferential statistics to calculate *d*-indexes (see Rosenthal, 1994).

For studies that involved naturalistic, cross-sectional measures of the amount of time spent on homework and related these to achievement but also included an attempt to statistically equate students on other characteristics, our preferred measure of relationship strength was the standardized beta-weight, β. These were derived either from the output of multiple regressions or as path coefficients in structural equation models. The standardized beta-weights indicate what change in the achievement measure expressed as a portion of a standard deviation was associated with a one-standard-deviation change in the homework variable. For example, if the standard deviation of the time-spent-on-homework variable equaled 1 hour and the standard deviation of the achievement measure equaled 50 points, then a beta-weight of .50 would mean that, on average, students in the sample who were separated by 1 hour of time-spent-on-homework also showed a 25-point separation on the achievement measure. In a few instances, beta-weights could not be obtained from study reports, so the most similar measures of effect (e.g., unstandardized regression weights, *b*) were retrieved. There were no instances in which we calculated beta-weights from other statistics.

For studies that involved naturalistic, cross-sectional measures but included no attempt to statistically equate students on third variables, we used simple bivariate correlations as measures of relationship. In some instances these were calculated from other inferential statistics (see Rosenthal, 1994).

Using three different measures of association implies that the relationship of homework to achievement cannot be compared across the three different types of design. This is not strictly true. Standardized mean differences and correlation coefficients can be transformed one to the other (see Cohen, 1988). A beta-weight equals a correlation coefficient when no other variables are controlled. However, we chose to present the results using each design's most natural metric so that the important distinction in their interpretation would not be lost.

Coder Reliability

Two coders extracted information from all reports selected for inclusion. Discrepancies were first noted and discussed by the coders, and if agreement was not reached the first author was consulted. Because all studies were independently coded twice and all disagreements resolved by a third independent coder, we did not calculate a reliability for this process (which would have entailed training three more coders and having them code at least a subset of studies).

Methods of Data Integration

Before conducting any statistical integration of the effect sizes, we first counted the number of positive and negative effects. For studies with effect size information, we calculated the median and range of estimated relationships. Also, we examined the distribution of sample sizes and effect sizes to determine if any studies contained statistical outliers. Grubbs's (1950) test, also called "the maximum normed residual test," was applied (see also Barnett & Lewis, 1994). This test identifies outliers in univariate distributions and does so one observation at a time. If outliers were identified, (using $p < .05$, two-tailed, as the significance level) these values would be set at the value of their next nearest neighbor.

Both published and unpublished studies were included in the synthesis. However, there is still the possibility that we did not obtain all studies that have investigated the relationship between homework and achievement. Therefore, we used Duval and Tweedie's (2000a, 2000b) trim-and-fill procedure to test whether the distribution of effect sizes used in the analyses were consistent with variation in effect sizes that would be predicted if the estimates were normally distributed. If the distribution of observed effect sizes was skewed, indicating a possible bias created either by the study retrieval procedures or by data censoring on the part of authors, the trim-and-fill method provides a way to estimate the values from missing studies that need to be present to approximate a normal distribution. Then, it imputes these missing values, permitting an examination of an estimate of the impact of data censoring on the observed distribution of effect sizes.

Calculating Average Effect Sizes

We used both weighted and unweighted procedures to calculate average effect sizes across all comparisons. In the unweighted procedure, each effect size was given equal weight in calculating the average value. In the weighted procedure, each independent effect size was first multiplied by the inverse of its variance. The sum of these products was then divided by the sum of the inverses. Generally speaking, weighted effect sizes are preferred because they give the most precise estimates of the underlying population values (see Shadish & Haddock, 1994). The unweighted effect sizes are also reported because in instances in which these are very different from the weighted estimates, this can give an indication that the magnitude of the effect size and sample size are correlated, sometimes suggesting that publication bias might be a concern. Also, 95% confidence intervals were calculated for weighted average effects. If the confidence interval did not contain zero, then the null hypothesis of no homework effect can be rejected.

Identifying Independent Hypothesis Tests

One problem that arises in calculating effect sizes involves deciding what constitutes an independent estimate of effect. Here, we used a shifting unit of analysis approach (Cooper, 1998). In this procedure, each effect size associated with a study is first coded as if it were an independent estimate of the relationship. For example, if a single sample of students permitted comparisons of homework's effect on both math and reading scores, two separate effect sizes were calculated. However, for estimating the overall effect of homework, these two effect sizes were averaged prior to entry into the analysis, so that the sample only contributed one effect size. To calculate the overall weighted mean and confidence interval, this one effect size would be weighted by the inverse of its variance (based primarily on sample size, which should be about equal for the two component effect sizes). However, in an analysis that examined the effect of homework on math and reading scores separately, this sample would contribute one effect size to each estimate of a category mean effect size.

The shifting unit of analysis approach retains as many data as possible from each study while holding to a minimum any violations of the assumption that data points are independent. Also, because effect sizes are weighted by sample size in the calculation of averages, a study with many independent samples containing just a few students will not have a larger impact on average effect size values than a study with only a single, or only a few, large independent samples.

Tests for Moderators of Effects

Possible moderators of homework–achievement relationships were tested by using homogeneity analyses (Cooper & Hedges, 1994; Hedges & Olkin, 1985). Homogeneity analyses compare the amount of variance in an observed set of effect sizes with the amount of variance that would be expected by sampling error alone. The analyses can be carried out to determine whether (a) the variance in a group of individual effect sizes varies more than predicted by sampling error, or (b) a group of average effect sizes varies more than predicted by sampling error. In the latter case, the strategy is analogous to testing for group mean differences in an analysis of variance or linear effects in a multiple regression.

Fixed and Random Error

When an effect size is said to be "fixed," the assumption is that sampling error is due solely to differences among participants in the study. However, it is also possible to view studies as containing other random influences, including differences in teachers, facilities, community economics, and so on. This view assumes that homework data from classrooms, schools, or even school districts in our meta-analysis also constitute a random sample drawn from a (vaguely defined) population of homework conditions. If it is believed that random variation in interventions is a significant component of error, a random-error model should be used that takes into account this study-level variance in effect sizes (see Hedges & Vevea, 1998, for a discussion of fixed and random effects).

Rather than opt for a single model of error, we chose to apply both models to our data. We conducted all our analyses twice, employing fixed-error assumptions once and random-error assumptions once. By employing this sensitivity analysis (Greenhouse & Iyengar, 1994), we could examine the effects of different assumptions on the outcomes of the synthesis. Differences in results based on which set of assumptions was used could then be part of our interpretation of results. For example, if an analysis reveals that a moderator variable is significant under fixed-error assumptions but not under random-error assumptions, this result suggests a limit on the generalizability of inferences about the moderator variable.

All statistical analyses were conducted using the Comprehensive Meta-Analysis statistical software package (Borenstein, Hedges, Higgins, & Rothstein, 2005).

Results

Studies With Exogenous Introductions of Homework

The literature search located six studies that employed a procedure in which the homework and no-homework conditions were imposed on students explicitly for the purpose of studying homework's effects. None of these studies was published. Some of the important characteristics and outcomes of each study are presented in Table 3.

Apparently, only one study used random assignment of students to conditions. McGrath (1992) looked at the effect of homework on the achievement of 94 high school seniors in three English classes studying the play Macbeth. At one point in the research report, the author states that half of the students "elected to receive no homework" and half "elected to receive homework" (p. 27). However, at another point, the report states that each student was assigned to a condition "by the alphabetic listing of his/her last name" (p. 29). Thus it might be (optimistically) assumed that the students in each of the three classes were haphazardly assigned to homework and no-homework conditions. In the analyses, the student was used as the unit. The experiment lasted 3 weeks and involved 12 homework assignments. Students doing homework did significantly better on a posttest achievement measure, $d = .39$.

A study by Foyle (1990) assigned four whole 5th-grade classrooms (not individual students) to conditions at random, one to a practice homework condition, one to a preparation homework condition, and two to a no-homework control condition. Clearly, assigning only one classroom to each condition, even when done at random, cannot remove confounded classroom differences from the effect of homework. For example, all four classrooms used a cooperative learning approach to teaching social studies, but one classroom (assigned to the practice homework condition) used a different cooperative learning approach from the other three classes. Also, the student, rather than the classroom, was used as the unit for statistical analysis, creating the concern that within-class dependencies among students were ignored. Analysis revealed that students differed significantly on a social studies pretest and on a standard measure of intelligence, but it was not reported whether there were preexisting classroom differences on these measures. Students doing homework outperformed no-homework students on unadjusted posttest scores, $d = .90$, and on posttest scores adjusted for pretest and intelligence differences, $d = .99$.

Foyle (1984) conducted a similar study on six high school classes in American history. Here, the experimenter reported that "the assignment of treatment and control groups was under the experimenter's control" (p. 90) and two intact classrooms were each assigned randomly to practice homework, preparation homework, and no-homework conditions. However, the student was again used as the unit of analysis. Analyses of covariance that controlled for pretest scores, aptitude differences, and the students' sex revealed that students doing homework had higher posttest achievement scores than students who did not. The covariance analysis and post hoc

TABLE 3
Studies that manipulated homework and no-homework conditions

Author and year	Research design	Number of classes and students, ESS	Grade level	Subject matter	Type of achievement measure	Effect size
Finstad, 1987	Nonequivalent control with no pretest differences	2 classes 39 students ESS = 5.2	2	Number place values to 100	Unit test developed by Harcourt Brace Jovanovich	+.97
Foyle, 1984	Randomized by class, analyzed by student	6 classes 131 students ESS = 15.8	9–12	American history	Unit test developed by the teacher	+.46
Foyle, 1990	Randomized by class, analyzed by student	4 classes 64 students ESS = 10.2	5	Social studies	Unit test developed by the teacher	+.90
McGrath, 1992	Randomized within class, analyzed by student	3 classes 94 students ESS = 8.0	12	Shakespeare	Unit test developed by Harcourt Brace Jovanovich	+.39
Meloy, 1987	Unlucky random assignment followed by nonequivalent control with pretest	5 classes 70 students ESS = 12.6	3	English skills	Unit test developed by McDugal, Little Researcher-shortened version of the Iowa Test of Basic Skills Language subtest	+ –
		3 classes 36 students ESS = 7.4	4		Unit test developed by McDugal, Little Researcher-shortened version of the Iowa Test of Basic Skills Language subtest	+ –
Townsend, 1995	Nonequivalent control without equating	2 classes 40 students ESS = 5.2	3	Vocabulary	Unit test developed by the teacher	+.71

Note. ESS = effective sample size based on an assumed intraclass correlation of .35.

tests revealed a significant positive effect of homework, but an effect size could not be calculated from the adjusted data (because the reported *F*-test contained two degrees of freedom in the numerator and means and standard deviations were not provided). The approximate, unadjusted homework effect was $d = .46$.

Finstad (1987) studied the effect of homework on mathematics achievement for 39 second-grade students in two intact classrooms. One unit, on place values to 100, was used, but neither the frequency nor the duration of assignments was reported. One classroom was assigned to do homework and the other not. It was not reported how the classroom assignments were carried out, but it was reported that there were no pretest differences between the classes. Data were analyzed on the student level without adjustment. The students in the classroom doing homework performed significantly better on a posttest measure, $d = .97$.

Meloy (1987) studied the effects of homework on the English skills (sentence components, writing) of third and fourth graders. Eight intact classrooms took part in the study and classes were matched on a shortened version of the Iowa Test of Basic Skills (ITBS) language subtest before entire classes were randomly assigned to homework and no-homework conditions. However, examination of pretest differences on the ITBS language subscale revealed that the students assigned to do homework scored significantly higher than students in no-homework classes. Thus a pretest-posttest design was used to control for the initial group differences, but pretests were used as a within-students factor rather than as a covariate (meaning a significant homework effect would appear as an interaction with time of testing). Also, students who scored above a threshold score on the pretest were excluded from the posttest analysis. Thus only 106 of an original sample consisting of 186 students were used in the analyses, and excluded students were not distributed equally across homework and no-homework conditions. Grade levels were analyzed separately, and classrooms were a factor in the analyses. The class-within-condition effect was not significant, so, again, the student was used as the unit of analysis. Homework was assigned daily for 40 instructional days. This study also monitored the homework completion rates in classrooms and set up reinforcement plans, different for each class, to improve completion rates. The effects of homework were gauged by using a researcher-modified version of the ITBS language subtest and a unit mastery test from the textbook. The complex reporting of statistical analyses made it impossible to retrieve simple effect estimates from the data. However, the author reported that the condition-by-time interactions indicated that homework had a significant negative effect on ITBS scores for third graders and a significant positive effect on fourth graders' unit test scores.

Finally, Townsend (1995) examined the effects of homework on the acquisition of vocabulary knowledge and understanding among 40 third-grade students in two classes, both taught by the experimenter. Treatment was given to classes as a whole and it was not stated how each class was assigned to the homework or no-homework condition. The student was used as the unit of analysis. A teacher-prepared posttest measure of vocabulary knowledge suggested that the homework group performed better, $d = .71$.

In sum, the six studies that employed exogenous manipulations all revealed a positive effect of homework on unit tests. One study (Meloy, 1987) revealed a negative effect on a standardized test modified by the experimenter. Four of the six studies employed random assignment, but in three cases assignment to conditions was carried out at the classroom level, using a small number of classrooms, and analyses were conducted using the student as the unit of analysis. In the only instance in which random assignment appears to have occurred within classes (McGrath, 1992), students also were used as the unit of analysis. Also, random assignment appears to have failed to produce equivalent groups in one study (Meloy, 1987).

While the introduction of homework as an exogenous intervention is a positive feature of these studies, other methodological considerations make it difficult to draw strong causal inferences from their results. Still the results are encouraging because of the consistency of findings. The measurable effects of homework on unit tests varied between $d = .39$ and $d = .97$. Also, the three studies that successfully used random assignment, fixed weighted $d = .53$ (95% CI = .29/.79), random weighted $d = .54$ (95% CI = .26/.82), produced effect sizes that were smaller than those of two studies that used other techniques to produce equivalent groups and for which effect sizes could be calculated, fixed weighted $d = .83$ (95% CI = .37/1.30), random weighted $d = .83$ (95% CI = .37/1.30); but the difference in mean d-indexes between these two sets of studies was not significant, fixed $Q(1) = 1.26$, *ns,* random $Q(1) = 1.12$, *ns.* Collapsing across the two study designs and using fixed-error assumptions, the weighted mean d-index across the five studies from which effect sizes could be obtained was $d = .60$ and was significantly different from zero (95% CI = .38/.82). Using a random-error model, the weighted average d-index was also .60 (95% CI = .38/.82).

To take into account the within-class dependencies that were not addressed in the reported data analyses, we recalculated the mean effect sizes and confidence intervals by using an assumed intraclass correlation of .35 to estimate effective sample sizes. In this analysis, the weighted mean d-index was .63, using both fixed and random-error assumptions, and both were statistically different from zero (95% CI = .03/1.23, for both). The mean d-index would not have been significant if an intraclass correlation of .4 was assumed. Additionally, the tests of the distribution of d-indexes revealed that we could not reject the hypothesis that the effects were estimating the same underlying population value when students were used as the unit of analysis, $Q_{fixed}(5) = 4.09$, *ns,* $Q_{random}(5) = 4.00$, *ns,* or when effective sample sizes were used as the unit, $Q_{fixed}(5) = .54$, *ns,* $Q_{random}(5) = .54$, *ns.*

And finally, the trim-and-fill analyses were conducted looking for asymmetry using both fixed and random-error models to impute the mean d-index (see Borenstein et al., 2005). Neither of the analyses produced results different from those described above. There was evidence that two effect sizes might have been missing. Imputing them would lower the mean d-index to $d = .48$ (95% CI = .22/.74) using both fixed and random-error assumptions.

The small number of studies and their variety of methods and contexts preclude their use in any formal analyses investigating possible influences on the magnitude of the homework effect, beyond comparing studies that used random assignment versus other means to create equivalent groups. The studies varied not only in research design but also in subject matter, grade level, duration, amount of homework, and the degree of alignment of the outcome measure with the content of assignments. Replications of any important feature that might influence the homework effect are generally confounded with other important features, and no visible pattern connecting effect sizes to study features is evident.

Studies Using Cross-Sectional Data and Control of Third Variables

Studies Using the National Education Longitudinal Study (1988, 1990, or 1992)

The literature search located nine reports that contained multivariate analyses of data collected as part of the National Education Longitudinal Study of 1988 (NELS) or in one of the NELS follow-ups on the same students in 1990, 1992, 1994, or 2000. These studies are described in Table 4. The NELS was conducted by the National Center for Educational Statistics and involved a nationally representative two-stage stratified probability sample. The final student sample in the first wave included 24,599 eighth-grade students. Each student completed achievement tests in mathematics, reading, science, and social studies in 1988, 1990, and 1992, as well as a 45-minute questionnaire that included questions about school, school grades, personal background, and school context. Various waves of the NELS also included surveys of teachers, school administrators, and parents. Student transcripts were collected at the end of their high school careers. Questions on homework were completed by both students and teachers, and they were asked about the total minutes of homework completed or assigned in different subject areas.

Several of the studies using the NELS data sampled students from the NELS itself for the purpose of examining questions regarding restricted populations. For example, Peng and Wright (1994) were interested in studying differences in relationships between predictors of achievement across ethnic groups, with a focus on Asian Americans. Davis and Jordan (1996) focused on African American males, while Roberts (2000) restricted the subsample to students attending urban schools only.

Examined as a group, the studies using NELS data use a wide variety of outcome measure configurations and different sets of predictor variables, in addition to homework. Still, every regression coefficient associated with homework was positive, and all but one were statistically different from zero. The exception occurred in the study of African American males on a composite measure of class grades (Davis & Jordan, 1996).

The study revealing the smallest beta-weight was a dissertation by Hill (2003). This report presents an unclear description of how the subsample drawn from the NELS was defined. The text reports that students were omitted from the sample if they "attended public schools, live in suburban areas, are neither Black nor Hispanic; and whose teachers are male, not certified in [the subject of the outcome variable], have neither an undergraduate degree in education or in [the subject of the outcome variable], and have neither a graduate degree in education or [the subject of the outcome variable]" (pp. 45, 86, 120). However, the tables in the report suggest that White students were included in the samples. The regression models suggest that students with teachers who had degrees in subjects other than the outcome variable also were included. Thus it is difficult to determine whether sampling restrictions might be the cause of the small regression coefficients associated with homework.

The dissertation by Lam (1996) deserves separate mention. In this study using data from 12th graders, the amount of homework students reported doing was entered into the regression equation as four dummy variables. This permitted an examination of possible curvilinear effects of homework. As Table 4 reveals, students who reported doing homework always had higher achievement scores than students who did not do homework (coded as the dummy variable). However, the strongest relationship between homework and achievement was found among students who reported doing

TABLE 4

Characteristics of studies using data from the National Education Longitudinal Study (1988, 1990, or 1992) and performing multivariate analyses

Author, year, and document type	Sample characteristics	Modeling technique	Outcome variable(s)	Predictor variables	Regression coefficient	Test result and significance
Davis & Jordan, 1994 Journal article	1,236 NELS: 88 Grade 8 African American males[a]	Multiple regression	Achievement test composite (math, science, and history achievement test scores)	Academic workload, urbanicity, attendance rate, school SES, number of Black teachers, discipline stressed, class size, college preparatory classes, dropout rate, teacher expectancies, motivation, teacher characteristics, homework, attendance, math coursework, science coursework, remediation, suspensions, retention, prior reading learning, SES	Achievement test composite: $\beta = .05$	$t = 2.21$, $p < .01$
			Class grades composite (self-reported grades in math, science, history, and English)		Class grades composite: $\beta = .03$	$t = 1.05$, $p > .05$
Hill, 2003 Analysis 1, dissertation	9,329 Math 1,104 Reading 12,302 Science NELS: 88, 90, 92 Grades 8, 10, 12[b]	Multiple regression (panel estimation)	Math Item Response Theory score	Homework (student report), school setting, SES, work time, TV time, teacher academic degree, class size, teaching style	Math: $\beta = .01$	$t = 28.85$, $p < .05$
			Reading Item Response Theory score		Reading: $\beta = .01$	$t = 13.96$, $p < .05$

	Multiple sample restrictions		Science Item Response Theory score		Science: $\beta = .01$	$t = 26.56$, $p < .05$
Hill, 2003 Analysis 2, dissertation	9,329 Math 1,104 Reading 12,302 Science NELS: 88, 90, 92 Grades 8, 10, 12[b] Multiple sample restrictions	Multiple regression (panel estimation)	Math Item Response Theory score	Homework (student report), school setting, SES, teacher ethnicity, teacher academic degree, class size, teaching style	Math: $\beta = .02$	$t = 31.31$, $p < .05$
			Reading Item Response Theory score		Reading: $\beta = .01$	$t = 16.00$, $p < .05$
			Science Item Response Theory score		Science: $\beta = .01$	$t = 29.42$, $p < .05$
Hill, 2003 Analysis 3, dissertation	9,329 Math 1,104 Reading 12,302 Science NELS: 88, 90, 92 Grades 8, 10, 12[b] Multiple restrictions	Multiple regression (panel estimation)	Math Item Response Theory score	Homework (teacher report), school setting, SES, TV time, work time, teacher ethnicity, teacher academic degree, class size, teaching style	Math: $\beta = .01$	$t = 11.62$, $p < .05$
			Reading Item Response Theory score		Reading: $\beta = .01$	$t = 8.73$, $p < .05$
			Science Item Response Theory score		Science: $\beta = .01$	$t = 12.59$, $p < .05$
Hill, 2003 Analysis 4, dissertation	9,329 Math 1,104 Reading 12,302 Science NELS: 88, 90, 92 Grades 8, 10, 12[b] Multiple restrictions	Multiple regression (panel estimation)	Math Item Response Theory score	Homework (teacher report), school setting, SES, teacher ethnicity, teacher academic degree, class size, teaching style	Math: $\beta = .01$	$t = 13.01$, $p < .05$
			Reading Item Response Theory score		Reading: $\beta = .01$	$t = 9.97$, $p < .05$
			Science Item Response Theory score		Science: $\beta = .01$	$t = 14.17$, $p < .05$

(continued)

TABLE 4 (*Continued*)

Author, year, and document type	Sample characteristics	Modeling technique	Outcome variable(s)	Predictor variables	Regression coefficient	Test result and significance
Lam, 1996 Dissertation	8,362 NELS: 92 Grade 12 175 Chinese Americans 8,187 Caucasian Americans	Multiple regression	Reading achievement test score	Ethnicity, sex, SES, TV time, absenteeism, hours of work per week, plus five dummy codes comparing students who do less than 1 hour of homework with those who do 1–6 hours of homework, 7–12 hours of homework, 13–20 hours of homework, more than 20 hours of homework	Reading: 1–6 hrs, $\beta = .13$ 7–12 hrs, $\beta = .21$ 13–20 hrs, $\beta = .16$ Over 20 hrs, $\beta = .12$	All $ps < .005$
			Math achievement test score		Math: 1–6 hrs, $\beta = .09$ 7–12 hrs, $\beta = .19$ 13–20 hrs, $\beta = .16$ Over 20 hrs, $\beta = .10$	 $p < .005$ $p < .05$ $p < .005$ $p < .005$
			Science achievement test score		Science: 1–6 hrs, $\beta = .09$ 7–12 hrs, $\beta = .17$ 13–20 hrs, $\beta = .12$ Over 20 hrs, $\beta = .09$	All $ps < .005$
			Social studies achievement test score		Social studies: 1–6 hrs, $\beta = .11$ 7–12 hrs, $\beta = .18$ 13–20 hrs, $\beta = .13$ Over 20 hrs, $\beta = .11$	All $ps < .005$
			Composite achievement test score		Composite: 1–6 hrs, $\beta = .12$ 7–12 hrs, $\beta = .21$ 13–20 hrs, $\beta = .16$ Over 20 hrs, $\beta = .12$	All $ps < .005$

Modi, Konstantopoulos, & Hedges, 1998 Conference paper	12,856 NELS: 92 Grade 12	Multiple regression	Achievement test composite (reading, math, science, and social studies)	Ethnicity, sex, educational aspirations, self-confidence, attendance, homework (more than 6 hours a week), reading time, TV time, religiosity, extracurricular activities, parent reliance, family background, location, school background	$\beta = .11$	$p < .0001$
Peng & Wright, 1994 Analysis 1, journal article	9,685[c] NELS: 88 Grade 8	Multiple regression	Achievement test composite (reading and math)	Family background, homework, TV time, extracurricular activities, educational aspirations, parental assistance with homework	$\beta = .10$	$p < .01$
Peng & Wright, 1994 Analysis 2, journal article	9,685[c] NELS: 88 Grade 8	Multiple regression	Achievement test composite (reading and math)	Ethnicity, family background, homework, TV time, extracurricular activities, educational aspirations, parental assistance with homework	$\beta = .08$	$p < .01$

(continued)

TABLE 4 (*Continued*)

Author, year, and document type	Sample characteristics	Modeling technique	Outcome variable(s)	Predictor variables	Regression coefficient	Test result and significance
Roberts, 2000 Dissertation	7,178 NELS: 88 Grade 8 Urban schools	Multilevel random coefficients modeling	Science achievement test score	Parent involvement, fear of asking questions in class, homework, non-English classroom	$b = 1.48$ Unweighted sample; multilevel sample	$p < .01$
		Multiple regression			$b = 1.10$ Weighted sample; multiple regression	$p < .01$
Schewior, 2001 Dissertation	7,138 NELS: 92 Grade 12	Multiple regression	Math achievement test score	Sex, ethnicity, location, school, family income, parent education level, teacher experience, class size, homework (two dummy codes comparing completion of homework "all of the time" with "most or all of the time")	"Most or all of the time," $b = 1.28$ "All of the time," $b = 2.59$	$t = 4.62$, $p < .0001$ $t = 8.09$, $p < .0001$
Thomas, 2001 Analysis 1, research report	450 NELS: 88 Grade 8 Picked at random	Logistic regression	Reading proficiency dichotomous score reflecting whether the	Sex, advanced math, grades, SES, homework, reading proficiency	$\beta = .28$	$p < .01$

			student exhibited proficiency at the highest level			
Thomas, 2001 Analysis 2, research report	450 NELS: 88 Grade 8 Picked at random	Logistic regression	Reading proficiency score recategorized to provide four exhaustive and nonoverlapping categories with higher scores indicating greater proficiency	School type, religiosity, sex, advanced math, ethnicity, grades, postsecondary plans, locus of control, self-concept, SES, parent education, homework, reading proficiency	$\beta = .17$	$p < .05$
Thomas, 2002 Research report	450 NELS: 88 Grade 8 Pick at random	Discriminant analysis[d]	Math proficiency score, reading proficiency score, science proficiency score (all 3-level scales)	Grades, motivation, advanced coursework, family background, homework, ethnicity, sex	Math: $\beta = .17$ Reading: $\beta = .13$ Science: $\beta = .15$	NR $F = 5.07$, $p = .02$ $F = 7.27$, $p = .02$

Note. NELS = National Education Longitudinal Study, NR = not reported.
[a]These students were required to have participated in NELS:90 as well.
[b]Students were included in the sample when data were available from NELS:88, NELS:90, and NELS:92; or from NELS:88 and NELS:90; or from NELS:88 and NELS:92.
[c]The authors write that this number represents the "effective" sample size, which was the actual sample size adjusted by the design effect, citing Ingels, Abraham, Karr, Spencer, & Frankel (1990).
[d]This study also reports results from canonical correlation and MANOVA analyses that are largely consistent with the findings of the discriminant analysis. The discriminant analysis is used here because it is most similar to analyses reported in other studies.

7 to 12 hours of homework per week, followed by students who reported doing 13–20 hours per week. Students who reported doing more than 20 hours of homework per week revealed a relationship with achievement test scores nearly equal to those reporting between 1–6 hours of homework per week. While this result is suggestive of a curvilinear relationship between homework and achievement, we must bear in mind that Lam restricted the sample of students to Asian Americans and Caucasian Americans.

In sum, if we omit (a) the Hill (2003) study (which produced beta-weights of .01 and .02), as well as (b) those studies that reported unstandardized regression weights, or (c) those for which coefficients could not be determined, then the reported beta-weights for the relation between homework and standardized achievement test scores range from .05 to .28. For composite achievement scores the range is from .05 to .21; for math, it is .09 to .16; for reading, .12 to .28; for science, .09 to .23; and for social studies, .11 to .18. Thus the ranges of estimated regression coefficients appear quite similar across the subject areas. However, we would caution against drawing any conclusions regarding the mediating role of subject matter on the homework–achievement relationship from these data, because the number and type of predictors in each model are confounded with subject matter. It should also be kept in mind that these estimates refer to high school students only.

Studies Using Data Other Than the National Education Longitudinal Study and Performing Multivariate Analyses

Table 5 provides information on 12 additional studies that performed multivariate analysis on cross-sectional data in order to examine the relationship between homework and achievement, with other variables controlled. Two of the studies used the High School and Beyond database (Cool & Keith, 1991; Fehrmann, Keith, & Reimers, 1987). The High School and Beyond database drew its 1980 base-year sample of sophomores and seniors from high schools throughout the United States. Probability sampling was used with overrepresentation of special populations. Follow-up surveys were conducted in 1982 and 1984. Brookhart (1997) used the Longitudinal Study of American Youth database, containing a national probability sample of approximately 6,000 seventh and tenth graders stratified by geographic area and degree of urban development. The rest of the studies used data collected by the researchers for the specific purpose of studying variables related to achievement.

Two studies conducted by Smith (1990, 1992), using overlapping data sets of seventh, ninth, and eleventh graders, found some negative relationships between homework and achievement. One of these findings (in Smith, 1992) revealed a small but statistically significant negative relationship between the amount of time spent on homework and language achievement, $\beta = -.06$. However, this study also revealed a significant positive interaction between year in school and time spent on homework. The interaction was not interpreted. This was the only significant negative result obtained in any of the cross-sectional, multivariate studies.

The remaining studies that used secondary school students all revealed positive and generally significant relationships. The three studies that used elementary school students (Cooper et al., 1998; Olson, 1988; Wynn, 1996) all revealed positive relationships between the homework measure and achievement (in Cooper et al., $\beta = .22$ for teacher-reported overall grades; in Olsen, $\beta = .10$ for math and $\beta = .11$

for reading; and in Wynn, $\beta = .04$ for grade point average). Thus, in addition to using varying predictor variables in the regression models, these studies also included a variety of outcome measures, including not only standardized tests but also teacher-assigned grades. In one instance, (Hendrix, Sederberg, & Miller, 1990) the outcome measure was not achievement but rather an indicator of school commitment/alienation constructed by the researcher that measured the importance of successful performance on school tasks, effort, and relevance of school work for student's lives. Thus we would again caution against drawing conclusions about mediating and moderating variables from these studies. It seems safest simply to note that the positive relationship between homework and achievement across the set of studies was generally robust across sample types, models, and outcome measures.

Structural Equation Modeling Studies Using Data From the National Education Longitudinal Study (1988, 1990, or 1992)

Table 6 provides information on four studies that tested structural equation models using data from the National Education Longitudinal Study. These analyses all revealed a positive relationship between the amount of time spent on homework and achievement. Not surprisingly, they are also somewhat larger than the relationships reported in studies that used multiple regression approaches to data analysis.

Structural Equation Modeling Studies Using Data From the High School and Beyond (1980, 1982, 1984) Longitudinal Studies

Table 7 provides information on four studies that tested structural equation models using data from the High School and Beyond database. All coefficients but one are positive and statistically significant. Keith and Benson (1992) found a nonsignificant negative coefficient for a subsample of Native Americans, $\beta = -.09$. The authors caution against strong interpretation of this finding because (a) the sample size was small ($n = 147$), and (b) Native American students who attended Bureau of Indian Affairs schools were not sampled. Still, it is generally the case that coefficients for the homework–achievement relationship estimated using High School and Beyond data are smaller than those estimated using NELS data.

Structural Equation Modeling Studies Using Original Data

We could find only one study that performed a structural equation analysis on data collected by the researchers. This study was also unique in that it examined the relationship between homework and achievement for elementary school students, a total of 214 second and fourth graders, who attended three adjacent school districts, one urban, one suburban, and one rural. Cooper, Jackson, Nye, and Lindsay (2001) used the MPlus program to predict grades assigned by teachers. In addition to the amount of homework that students reported doing, the model included student ability and homework norms, parent attitude, home environment (e.g., TV and quiet time), parent facilitation, presence of alternative activities, and student attitudes. The path coefficient for the relationship between time on homework and class grade was .20, $p < .01$.

Studies Correlating Time on Homework and Academic Achievement

The literature search uncovered 32 studies that described the correlations between the time that a student spent on homework, as reported by either the student or a

TABLE 5

Characteristics of studies using data other than those from the National Education Longitudinal Study (1988, 1990, or 1992) and performing multivariate analyses

Author, year, and document type	Sample characteristics	Modeling technique	Outcome variable(s)	Predictor variables	Regression coefficient	Test result and significance
Brookhart, 1997 Journal article	118[a] Longitudinal Study of American Youth, 1987–1991 Grade 12	Path analysis: stepwise multiple regression	Math achievement test score	Sex, SES, reading comprehension, prior Math achievement (Grades 10–11), classroom assessment environment (% class time testing; Grades 10–12), hours of homework assigned, % completing homework on time, % of homework returned	$\beta = .09$	$p < .05$
Cool & Keith, 1991 Journal article	28,051, with pairwise deletion, minimum $N = 21{,}008$ High School and Beyond Longitudinal Study, 1980 Grade 12	Path analysis	Achievement test composite (reading, math I and II scores)	Exogenous: ethnicity, sex, family background Endogenous: ability, quality of instruction, motivation, homework, quantity of coursework	$\beta = .04$	*ns*
Cooper, Lindsay, Nye, & Greathouse, 1998	285 Original data Grades 2, 4	Path analysis: multiple regression	Teacher-reported class grades	Exogenous: standardized achievement test score, parent attitudes, teacher attitudes	$\beta = .13$	$p < .02$

Analysis 1, journal article				Endogenous: amount of homework assigned, student attitudes, homework completion		
Cooper, Lindsay, Nye, & Greathouse, 1998 Analysis 2, journal article	424 Original data Grades 6–12	Path analysis: multiple regression	Teacher-reported class grades	Exogenous: standardized achievement test score, parent attitudes, teacher attitudes Endogenous: amount of homework assigned, student attitudes, homework completion	$\beta = .22$	$p < .0001$
Fehrmann, Keith, & Reimers, 1987 Journal article	28,051 High School and Beyond Longitudinal Study, 1980 Grade 12	Path analysis: multiple regression	Student-reported class grades	Exogenous: ethnicity, family background, sex Endogenous: ability, parental involvement, homework, TV time	$\beta = .19$	NR
Foyle, 1984 Dissertation	131 Original data Grade 10	ANCOVA	American history posttest scores	Covariates: pretest score, ability Main effects: sex, homework (practice, preparatory, or no homework) Interaction: Sex × homework	Not obtainable	$F = 7.81$, $p < .001$

(continued)

TABLE 5 (*Continued*)

Author, year, and document type	Sample characteristics	Modeling technique	Outcome variable(s)	Predictor variables	Regression coefficient	Test result and significance
Hendrix, Sederberg, & Miller, 1990 Journal article	1,521 Original data Grade 12	Stepwise regression	School commitment/alienation	Nonverbal ability, verbal ability, % advanced courses, TV time, homework, job time, GPA; nonverbal ability × TV; verbal ability × GPA; advanced courses × GPA	$\beta = .21$	$t = 8.91$, $p < .001$
Olson, 1988 Dissertation	191 Original data Grades 3–6	Multiple regression	Math achievement test score (California Test of Basic Skills), reading achievement test score (California Test of Basic Skills)	Homework achievement, homework, student homework behavior, parent-involvement, % preparatory homework, % distributed homework, homework feedback, student ability	Math: $\beta = .10$ Reading: $\beta = .11$	$F = 2.71$, *ns* $F = 3.07$, *ns*
Portes & MacLeod, 1996 Journal article	5,266[b] Original data Grades 8, 9	Multiple regression	Math achievement test score (Stanford Achievement Test), reading achievement test score (Stanford Achievement Test)	Age, sex, parental SES, length of U.S. residence, homework, region, ethnicity	Math: $b = 2.50$ Reading: $b = 1.74$	$p < .01$ $p < .01$

Portes & Zady, 2001 Conference paper	5,264 Same data as used by Portes & MacLeod, 1996 Grades 8, 9	Hierarchical multiple regression	Reading achievement test score (Stanford Achievement Test)	Block 1: grade, age, English proficiency, SES, sex, school location, length of stay in U.S. Block 2: economic situation, homework, TV time, peer relations, familialism, perceived English proficiency, discrimination, motivation, ethnic pull, father presence, self-esteem	Spanish speakers: $\beta = .07$ Asian origin: $\beta = .06$	$p = .002$ $p = .03$
Smith, 1990 Journal article	1,584 Original data Grade 7, 9	Multiple regression	Overall achievement test score (California Test of Basic Skills) Reading scores Language scores Math scores	Ethnicity, sex, year in school, family background, chore time, ethnicity × chore time, homework, friends, year × friends, TV time, job × TV time, radio and recordings time, reading time	Overall: $\beta = .00$ Reading: $\beta = -.02$ Language: $\beta = .00$ Math: $\beta = .04$	*ns* *ns* *ns* *ns*

(continued)

TABLE 5 (*Continued*)

Author, year, and document type	Sample characteristics	Modeling technique	Outcome variable(s)	Predictor variables	Regression coefficient	Test result and significance
Smith, 1992 Journal article	1,208 Original data[c] Grades 9, 11	Multiple regression	Overall achievement test scores (California Test of Basic Skills) Reading scores Language scores Math scores	Previous achievement, ethnicity, sex, year in school, family background, parent time, chore time, homework, year × homework, friends, TV time, occupation × TV time, radio and recordings time, reading time	Overall: $\beta = -.02$ Reading: $\beta = -.02$ Language: $\beta = -.06$ Math: $\beta = .01$	*ns* *ns* $p < .01$ *ns*
Wynn, 1996 Dissertation	170 Original data Grade 3	Path analysis: multiple regression	GPA (combined reading, math, spelling, language arts, social studies, and science)	Exogenous: ethnicity, SES, sex Endogenous: ability, family involvement, homework, TV time	$\beta = .04$	*ns*

Note. NR = not reported.

[a]This is a weighted sample size for 12th-grade students who participated in LSAY in all 3 years from 1987 to 1989 and who completed the mathematics achievement test.

[b]This sample consisted of Cuban, Vietnamese, Haitian, and Mexican second-generation children of immigrants from Florida and California.

[c]For this study, achievement test data from 1992 were added to the Smith (1990) data set. Thus, data are from the same students as in Smith (1990).

TABLE 6

Characteristics of structural equation modeling studies using data from the National Education Longitudinal Study (1988, 1990, or 1992)

Author, year, and document type	Sample characteristics	Program used	Outcome variable(s)	Predictor variables	Regression coefficient	Test result and significance
Bruce & Singh, 1996 Journal article	24,599 with pair-wise deletion $N = 21{,}835$, minimum $N = 18{,}029$ NELS: 88; Grade 8	LISREL	Achievement test composite (math, science, reading, and social studies)	Exogenous: ethnicity, family background, sex Endogenous: previous achievement, quality of instruction, motivation, homework, quantity of instruction	$\beta = .06$	NR
Keith, Diamond-Hallam, & Fine, 2004 Analysis 1, journal article	13,546 NELS: 88, 90, 92; Grades 8, 10, 12	Amos	1990 High school grades (English, math, science, and social studies) 1992 Achievement test scores (reading, math, science, and social studies)	Exogenous: ethnicity, family background Endogenous: Grade 8 achievement, homework in Grades 10 and 12 (in-school), homework in Grades 10 and 12 (out-of-school)	Class grades: $\beta = .28$ Achievement test scores: $\beta = .13$	NR NR
Keith, Diamond-Hallam, & Fine, 2004[a] Analysis 2, journal article	13,546 NELS: 88, 90, 92; Grades 8, 10, 12	Amos	1992 High school grades (English, math, science, and social studies)	Exogenous: ethnicity, family background Endogenous: Grade 8 achievement, homework in Grade 10 (in-school), homework in Grade 10 (out-of-school)	$\beta = .18$	NR
Keith & Keith, 1993[a] Journal article	21,814, with pair-wise deletion Minimum $N = 18{,}355$ NELS: 88; Grade 8	LISREL	Achievement test scores (reading, math, science, and social studies)	Exogenous: ethnicity, family background Endogenous: Previous achievement, parental involvement, homework, TV time	$\beta = .19$	NR
Singh, Granville, & Dika, 2002 Journal article	3,227 NELS: 88; Grade 8[b]	LISREL	Math achievement (class grades in math, Grades 6 and 8, math achievement test score)	Exogenous: attendance motivation, coursework motivation Endogenous: math attitude, academic time (math homework, TV time)	$\beta = .50$[c]	$p < .05$
			Science achievement (class grades in science, Grades 6 and 8, science achievement test score)	Exogenous: attendance motivation, coursework motivation Endogenous: science attitude, academic time (science homework, TV time)	$\beta = .61$[c]	$p < .05$

Note. NELS = National Education Longitudinal Study. [a]These data also appear in a conference paper by Keith, Keith, Bickley, and Singh (1992). [b]The data were a 25% random sample from the NELS:88 data set. [c]A "clean" coefficient from homework to mathematics achievement could not be obtained for this path. The reported direct path coefficient includes parameters associated with TV time combined with time spent on mathematics homework.

TABLE 7
Characteristics of structural equation modeling studies using data from the High School and Beyond (1980, 1982, or 1984) longitudinal studies

Author, year, and document type	Sample characteristics	Program used	Outcome variable(s)	Predictor variables	Regression coefficient	Test result and significance
Camp, 1990 Journal article	7,668 Grade 10	LISREL	Student-reported class grades, high school GPA	Exogenous: sex, academic ability, family background Endogenous: TV time, homework, job time, student participation in activities	$\beta = .06$	$t = 5.46$, $p < .05$
Keith & Benson, 1992 Journal article	13,152, with pairwise deletion $N = 12{,}142$, minimum $N = 8{,}910$ 1980, 1982, 1984 Grades 10, 12	LISREL	High school grades composite	Exogenous: ethnicity, sex, family background Endogenous: ability, quality of instruction, motivation, homework, New Basics coursework	$\beta = .06$	NR
Keith & Benson, 1992 Journal article	5,658 Caucasian, 1,039 African American, 1,791 Hispanic, 248 Asian American, 147 Native American (minimum *N*s) 1980, 1982, 1984 Grades 10, 12	LISREL	High school grades composite	Ability, quality of instruction, motivation, homework, New Basics coursework	Caucasian: $\beta = .06$ African American: $\beta = .09$ Hispanic: $\beta = .05$ Asian American: $\beta = .25$ Native American: $\beta = -.09$	$p < .05$ $p < .05$ $p < .05$ $p < .05$ *ns*
Keith & Cool, 1992 Journal article	25,875, with pairwise deletion Minimum $N = 21{,}427$ 1980, 1982 Grades 10, 12	LISREL	1982 Achievement test composite (combined reading, math I and II, science, writing, civics)	Exogenous: ethnicity, sex, family background Endogenous: ability, quality of instruction, motivation, quantity of coursework, homework	$\beta = .06$	$p < .05$

parent, and a measure of academic achievement. These studies are listed in Table 8. The 32 studies reported 69 separate correlations based on 35 separate samples of students. Cooper et al. (1998) reported 8 correlations, separating out effects for elementary and secondary students (two independent samples) on both class grades and standardized tests with time on homework reported by either students or parents. Drazen (1992) reported 12 correlations, for reading, math, and multiple subjects for three national surveys (three independent samples). Bents-Hill and colleagues (1988) reported 8 correlations, for language arts, math, reading, and multiple subjects both for class grades and for a standardized test of achievement. Epstein (1988), Olson (1988), and Walker (2002) each reported 2 effect sizes, 1 for math and 1 for reading. Fehrmann et al. (1992), Wynn (1996), and Keith and Benson (1992) each reported 2 correlations, 1 involving class grades and 1 involving achievement test results. Hendrix et al. (1990) reported 3 correlations, 1 for multiple subjects, 1 for verbal ability, and 1 for nonverbal ability. Mau & Lynn (2000) reported 3 correlations, 1 for math, 1 for reading, and 1 for science. Singh et al. (2002) reported 2 correlations for math and 1 for science.

The 32 studies appeared between the years 1987 and 2004. The sample sizes ranged from 55 to approximately 58,000 with a median size of 1,584. The mean sample size was 8,598 with a standard deviation of 12,856, suggesting a nonnormal distribution. The Grubbs test revealed a significant outlier, $p < .05$. This sample was the largest in the data set, reported by Drazen (1992) for six correlations obtained from the 1980 High School and Beyond longitudinal study. As a result, we replaced these six sample sizes with the next largest sample size in the data set, 28,051. The mean sample size for the adjusted data set was 7,742 with a standard deviation of 10,192.

Only three studies specifically mentioned that students were drawn from regular education classrooms, and one of these studies included learning-disabled students as well (Deslandes, 1999). The remaining studies did not report information on the students' achievement or ability level. Seventeen studies did not report information on the socioeconomic status of students, 11 reported that the sample's SES was "mixed," 3 described the sample as middle SES, and 1 as lower SES. Seventeen studies did not report the sex make-up of the sample, while 14 reports said the sample was comprised of both sexes. Only one study reported correlations separately for males and females. Because of a lack of reporting or variation across categories, no analyses were conducted on these variables.[1]

Of the 69 correlations, 50 were in a positive direction and 19 in a negative direction. The mean unweighted correlation across the 35 samples (averaging multiple correlations within each sample) was $r = .14$, the median was $r = .17$, and the correlations ranged from −.25 to .65.

The weighted average correlation was $r = .24$ using a fixed-error model with a 95% confidence interval (95% CI) from .24 to .25. The weighted average correlation was $r = .16$ using a random-error model with a 95% confidence interval from .13 to .19. Clearly, then, the hypothesis that the relationship between homework and achievement is $r = 0$ can be rejected under either error model. There were no significant outliers among the correlations, so all were retained for further analysis.

The trim-and-fill analyses were conducted in several different ways. We performed the analyses looking for asymmetry, using both fixed and random-error models to impute the mean correlation and creating graphs using both fixed and random

TABLE 8
Characteristics of studies correlating time on homework and academic achievement

Author and year	Document type	Number of students	Respondent type	Outcome measure	Grade level	Subject matter	Correlation
Antonek, 1996	Unpublished	89	Students	Other test	3–5	Foreign language	+.26
Bents-Hill, 1988	Unpublished	1,865	Parents	Class grades	3, 6	Language arts	−.01
						Math	−.04
						Reading	−.04
						Multiple subjects	−.03
						Language arts	−.06
				Standardized test		Math	−.08
						Reading	−.07
						Multiple subjects	−.09
Bowen & Bowen, 1998	Published	538	Students	Class grades and relative standing	Middle and high school	Multiple subjects	+.20
Broxie, 1987	Unpublished	55	Students	Class grades	4–6	Multiple subjects	+.65
Bruce, 1996	Published	21,835	Students	Standardized test	8	Multiple subjects	+.20
Cool & Keith, 1991	Published	28,051	Students	Standardized test	12	Multiple subjects	+.30
Cooper, Lindsay, Nye, & Greathouse, 1998	Published	~285	Students	Class grades,	2, 4	Multiple subjects	−.19
				standardized test			−.04
			Parents	Class grades,			−.13
				standardized test			−.06
		~424	Students	Class grades,	6–12		+.17
				standardized test			.00
			Parents	Class grades,			+.24
				standardized test			+.14
Deslandes, 1999	Published	637	Students	Class grades	High school	Language arts	+.18[a]

Drazen, 1992	Unpublished	~19,000 (NELS: 72)	Students	Standardized test	12	Reading	+.17
						Math	+.20
						Multiple subjects	+.20
		~58,000 (High School and Beyond: 80)[b]			10	Reading	+.25
						Math	+.29
						Multiple subjects	+.30
					12	Reading	+.23
						Math	+.28
						Multiple subjects	+.27
		~25,000 (NELS: 88)			8	Reading	+.17
						Math	+.20
						Multiple subjects	+.20
Epstein, 1988	Unpublished	1,021	Parents	NR	1, 3, 5	Math	−.05
						Reading	−.11
Fehrman, Keith, & Reimers, 1987	Published	28,051	Students	Class grades, standardized test	12	Multiple subjects	+.32
							+.25
Hendrix, Sederberg, & Miller, 1990	Published	1,521	Students	Class grades	12	Multiple subjects	+.35
						Nonverbal ability	+.16
				NR		Verbal ability	+.17
Hightower, 1991	Unpublished	9,002	Students	Standardized test	12	Multiple subjects	+.29
Keith & Benson, 1992	Published	8,910	Students	Class grades, standardized test	10, 12	Multiple subjects	+.30
							+.22
Keith & Cool, 1992	Published	21,427	Students	Standardized test	10, 12	Multiple subjects	+.30
Keith, Diamond-Hallam, & Fine, 2004	Published	6,773	Students	Standardized test	12	Multiple subjects	+.22
Lam, 1996	Unpublished	3,657	Students	Standardized test	12	Multiple subjects	+.04
Mau & Lynn, 2000	Published	20,612	Students	Standardized test	10, 12	Math	+.29
						Reading	+.24
						Science	+.23
Olson, 1988	Unpublished	191	Students	Standardized test	3–6	Math	+.11
						Reading	+.10

(continued)

TABLE 8 (*Continued*)

Author and year	Document type	Number of students	Respondent type	Outcome measure	Grade level	Subject matter	Correlation
Peng & Wright, 1994	Published	24,599	Students	Standardized test	8	Multiple subjects	+.17
Pezdek, Berry, & Renno, 2002	Published	380	Parents	Another test	4–6	Math	+.15[c]
Roberts, 2000	Unpublished	7,178	Students	Standardized test	8	Science	+.26
Rozevink, 1995	Unpublished	363	Students	Standardized test	9, 12	Multiple subjects	−.23
Schewior, 1992	Unpublished	4,930	Students	Standardized test	12	Math	+.20[d]
Singh, Granville, & Dika, 2002	Published	3,227	Students	Class grades	8	Math	+.11
						Science	+.10
				Standardized test		Math	+.30
Smit, 1990	Published	1,584	Students	Standardized test	7, 9	Multiple subjects	−.08
Thomas, 2001	Unpublished	450	Students	Standardized test	8	Math	+.22
Tonglet, 2000	Unpublished	189	Students	Class grades	5, 8	Math	+.47
Vazsonyi & Pickering, 2003	Published	764	Students	Class grades	High school	Multiple subjects	−.03
Walker, 2002	Unpublished	86	Students	Standardized test	4, 5	Math	+.17
						Reading	+.25
Wynn, 1996	Unpublished	170	Parents	Class grades,	3	Multiple subjects	+.00
				standardized test			−.17
Wynstra, 1995	Unpublished	68	Parents	Standardized test	1–5	Language arts	−.25

Note. NELS = National Education Longitudinal Study, NR = no response.
[a]This effect size is collapsed across general classroom ($n = 525$, $r = .17$) and learning disabled ($n = 112$, $r = .20$) students.
[b]Effect sizes were also presented for High School and Beyond: 82: Reading, $r = .26$; Math, $r = .28$; multiple subjects, $r = .29$. These effect sizes were not used in the primary analyses because they represent the same students who were surveyed in High School and Beyond: 80.
[c]This effect size was computed from correlations between the amount of time that students spent on mathematics homework and mathematics achievement across two studies for which grade level was statistically controlled post hoc (Study 1, $n = 165$, $r = .16$; Study 2, $n = 215$, $r = .14$).
[d]This effect size is collapsed across two effect sizes representing two levels of student-reported time spent on homework. Students who completed homework all of the time ($r = +.27$) reported a larger effect for homework than did students who completed their homework most or all of the time ($r = +.13$).

models (see Borenstein et al., 2005) while searching for possible missing correlations on the left side of the distribution (those that would reduce the size of the positive correlation). None of the analyses produced results different from those described above. When we used a random-error model, there was evidence that three effect sizes might have been missing and that imputing them would lower the mean fixed-effect correlation to $r = .23$ (95% CI = .22/.23). The random-error results of this analysis were $r = .14$ (95% CI = .11/.17).[2]

Next, we carried out a moderator analysis examining the association between the magnitude of correlations and the publication status of the study report. Seventeen of the samples had been published and their results were compared with those of the 18 samples that had appeared as dissertations, ERIC documents, or unpublished research reports. Under the fixed-error model, correlations from journal articles, $r = .25$, were significantly higher than those from unpublished sources, $r = .23$, $Q(1) = 20.71$, $p < .0001$. Under the random-error model, correlations from journal articles, $r = 18$, were not statistically different from those from unpublished sources, $r = .15$, $Q(1) = 0.91$, *ns*. In both instances, the absolute size of the difference was quite small.

Moderator Analyses

Table 9 presents the results of analyses examining whether the magnitude of the correlation between time spent on homework and achievement was moderated by the type of achievement measure. Two studies using unstandardized tests scores, one using a composite of standardized tests and class grades, and one not reporting the type of achievement outcome were omitted from this analysis because there were too few studies in each of these outcome-type categories. Thus the moderator analysis compared results involving class grades with results involving standardized achievement tests.

Under fixed-error assumptions, the correlation between time spent on homework and class grades, $r = .27$ (95% CI = .26/.27), was significantly higher than that involving standardized achievement test scores, $r = .24$ (95% CI = .24/.25), $Q(1) = 26.26$, $p < .0001$. Under random-error assumptions, the correlation between time spent on homework and class grades, $r = .19$ (95% CI = .11/.27), was not significantly different from that involving standardized achievement test scores, $r = .16$ (95% CI = .14/.19), $Q(1) = 0.35$, *ns*. In both instances, the absolute difference between the correlations was quite small.

Table 9 also presents the results of analyses examining whether the magnitude of the correlation between time spent on homework and achievement was moderated by the grade level of the students. Correlations were grouped into those involving elementary school students, Grades K–6, and secondary school students, Grades 7–12. One study (Tonglet, 2000) was omitted from the analysis because it included students in Grades 5 and 8 and the correlation for the two grades could not be separated. One correlation from Cooper et al. (1998) was omitted from the analysis because it included students in Grades 6–12. Tonglet (2000) and Cooper et al. (1998) reported sampling students from Grades 5 and 8, and 6–12, respectively. The correlation between time spent on homework and class grades was +.47 for Tonglet. The correlation was +.21 for Cooper et al., who also reported a correlation of +.07 between time spent on homework and standardized achievement test scores.

TABLE 9

Results of moderator analyses involving correlations between time on homework and measures of academic achievement

Grouping variable	Number of comparisons	95% Confidence interval: Low estimate		Mean		High estimate	
Overall	35	+.240	(+.134)	+.243	(+.161)	+.246	(+.188)
Outcome:							
$Q(1) = 26.26^{*}$ (.35)							
Class grades	12	+.26	(+.11)	+.27*	(+.19*)	+.27	(+.27)
Standardized test	26	+.24	(+.14)	+.24*	(+.16*)	+.25	(+.19)
Grade level:							
$Q(1) = 710.68^{*}$ (10.43*)							
K–6	10	−.06	(−.03)	−.04*	(+.05)	−.02	(+.13)
7–12	23	+.25	(+.17)	+.25*	(+.20*)	+.25	(+.22)
Subject matter:							
$Q(2) = 164.62^{*}$ (2.46)							
Reading	8	+.20	(+.07)	+.21*	(+.12*)	+.21	(+.18)
Math	13	+.24	(+.13)	+.24*	(+.18*)	+.25	(+.22)
Multiple subjects	18	+.24	(+.12)	+.25*	(+.16*)	+.25	(+.20)
Respondent:							
$Q(1) = 631.70^{*}$ (20.06*)							
Students	30	+.25	(+.16)	+.25*	(+.19*)	+.25	(+.21)
Parents	7	−.05	(−.10)	−.03*	(−.02)	−.01	(+.07)

Note. $^{*}p < .05$. Random effects Q_b values and point estimates are presented in parentheses.

Figure 1 presents a stem-and-leaf display of the 33 correlations associated with this analysis.

Under fixed-error assumptions, the correlation between time spent on homework and achievement was significantly higher for secondary school students, $r = .25$ (95% CI = .25/.25), than for elementary school students, $r = -.04$ (95% CI = −.06/−.02), $Q(1) = 710.68$, $p < .0001$. Under random-error assumptions, the correlation between time spent on homework and achievement was also significantly higher for secondary school students, $r = .20$ (95% CI = .17/.22), than for elementary school students, $r = .05$ (95% CI = −.03/.13), $Q(1) = 10.43$, $p < .002$. As indicated by the confidence intervals, using the random-error model, the mean correlation between time spent on homework and achievement was not significantly different from zero for elementary school students.

Table 9 also presents the results of analyses examining whether the homework–achievement correlation was moderated by the subject matter of the homework assignment. One study involving science, 1 involving foreign language, and 1 involving verbal and nonverbal ability were omitted from the analysis because there were too few studies in each of these outcome-type categories. Thus the moderator analysis compared only studies involving language arts, reading, mathematics, and achievement across multiple subject domains.

First, we compared correlations involving language arts with correlations involving reading. Using fixed-error assumptions, the three correlations involving language arts revealed a nonsignificant average weighted correlation of $r = -.01$ (CI = −.04/.02), while the eight reading outcomes produced a significant positive correlation of $r = .21$ (CI = .20/.21). These average correlations were significantly different from one another, $Q(1) = 202.94$, $p < .0001$. Using random-error assumptions, the average language arts correlation was nonsignificant, $r = .01$ (CI = −.10/.13), while reading produced a significant positive correlation, $r = .12$ (CI = .07/.18). These average

Lower Grades (1–6)	Stem	Upper Grades (7–12)
5	+.6	
	+.3	00
6	+.2	998665
1	+.2	32200000
5	+.1	877
1	+.1	
	+.0	4
689	−.0	38
1	−.1	
5	−.2	3

FIGURE 1. *Distribution of correlations between time on homework and achievement as a function of grade level.*

correlations approached being significantly different from one another, $Q(1) = 2.71$, $p < .10$. Because of these results, we chose not to combine the language arts and reading data sets but instead to use only reading correlations in the subsequent analyses examining subject matter as a moderator.

The average weighted correlations between time on homework and reading, math, and multiple subjects were significantly different from one another under fixed-error assumptions, $Q(2) = 164.62$, but not under random-error assumptions, $Q(2) = 2.46$, *ns.* We then proceeded to conduct two planned comparisons, one comparing reading outcomes with math outcomes and one comparing both math and reading outcomes with outcomes involving measures of multiple subjects.

Under fixed-error assumptions, the correlation between time spent on homework and achievement was significantly higher for math, $r = .24$ (95% CI = .24/.25) than for reading, $r = .21$ (95% CI = .20/.21), $Q(1) = 99.92$, $p < .0001$. Under random-error assumptions, the correlation between time spent on homework and achievement was not significantly different for math, $r = .18$ (95% CI = .13/.23), than for reading, $r = .12$ (95% CI = .07/.18), $Q(1) = 2.46$, *ns.* In both instances, the absolute difference between the correlations was quite small.

Under fixed-error assumptions, the correlation between time spent on homework and achievement was significantly higher for multiple subjects, $r = .25$ (95% CI = .25/.25) than for either reading or math alone, $r = .23$ (95% CI = .22/.23), $Q(1) = 64.70$, $p < .0001$. Under random-error assumptions, the correlation between time spent on homework and achievement was not significantly different for multiple subjects, $r = .16$ (95% CI = .12/.20), in comparison with that for reading or math alone, $r = .16$ (95% CI = .12/.19), $Q(1) = 0.004$, *ns.* Again, in both instances, the absolute difference between the correlations was quite small.

Finally, Table 9 presents the results of analyses examining whether the homework and achievement correlation was moderated by who provided data on the amount of time spent on homework. All studies included information about whether it was the student or a parent who was the respondent.

Under fixed-error assumptions, the correlation between time spent on homework and achievement was significantly higher when students made the report, $r = .25$ (95% CI = .25/.25) than when parents reported, $r = -.03$ (95% CI = −.05/−.01), $Q(1) = 631.70$, $p < .0001$. Under random-error assumptions, the correlation between time spent on homework and achievement was still significantly stronger for students, $r = .19$ (95% CI = .16/.21), than for parents, $r = -.02$ (95% CI = −.10/.07), $Q(1) = 20.06$, $p < .0001$. Using the random-error models, the correlations involving parent reports were not significantly different from zero.

Tests for Interactions Among Moderators

We next tested whether the main effects of moderator variables also held when tested within levels of other moderator variables. Specifically, we tested (a) whether the grade level of the student was associated with the magnitude of the homework–achievement correlation when the student was tested within different types of outcome measures; (b) whether the grade level of the student was associated with the magnitude of correlation when the student was tested within different types of subject matter; and (c) whether the subject matter of homework was associated with the magnitude of the homework–achievement correlation when the student was tested within different types of outcome measures.

The findings produced a pattern of results regarding the direction and significance for the moderator's effect that was consistent with the main effects in 13 of the 14 subgroup analyses. That is, both the direction of the comparison between correlations and the significance of the difference between correlations (using both fixed and random models) was the same when we compared the subgroup analyses to the main effect analyses in all instances but one. The exception was that when we used a random-error model to compare the relationship between homework and class grades for four correlations at the elementary school level, $r = .09$ (95% CI = −.10/.28), and six correlations at the secondary level, $r = .21$ (95% CI = .12/.30), the difference was not significantly different from zero, $Q(1) = 1.18$, *ns*. The direction of the difference between the mean correlations was the same as that in the main effect analyses.

Finally, we looked to see whether the respondent providing information about homework (the student or a parent) was confounded with any of the other three moderator variables. We found that 3 times parents provided information on homework in correlations involving class grades and 4 times when correlations involved achievement tests. Similarly, 3 times parents provided information when homework was associated with math, 2 times when associated with reading, and 3 times with multiple subjects.

However, all parent reports on the amount of homework were provided for students who were in Grades K–6.[3] Therefore it was possible that the significant difference suggesting that the homework–achievement relationship was smaller for elementary school than secondary school students might not hold if students were respondents. To test this hypothesis, we re-ran the grade level analyses using only students as respondents.

Under fixed-error assumptions, the correlation between time spent on homework and achievement was significantly higher for secondary school students, $r = .25$ (95% CI = .25/.25), than for elementary school students, $r = .06$ (95% CI = −.00/.11), $Q(1) = 47.48$, $p < .0001$. Under random-error assumptions, the correlation between time spent on homework and achievement was not significantly higher for secondary school students, $r = .19$ (95% CI = .17/.22), than for elementary school students, $r = .22$ (95% CI = .00/.42), $Q(1) = 0.57$, *ns*.

In light of these results, it is not surprising that we also found differences between student and parent reports at the elementary school level. Under fixed-error assumptions, the correlation between time spent on homework and achievement was significantly higher when elementary school students made the report, $r = .06$ (95% CI = .00/.11), than when parents of elementary school students made the report, $r = -.06$ (95% CI = −.08/−.04), $Q(1) = 14.40$, $p < .001$. Under random-error assumptions, the correlation between time spent on homework and achievement was still significantly stronger for elementary school student reports, $r = .22$ (95% CI = −.00/.42), than for parents, $r = -.05$ (95% CI = −.11/.01), $Q(1) = 5.40$, $p < .03$. It appears that, for elementary school students, parents report a small negative relationship between the amount of time their child spends on homework and their achievement, while the students themselves report a positive relationship.

Studies Correlating Time on Homework and Non-Achievement Measures

We found 5 studies that presented correlations between the amount of time students spent doing homework and student attitudes. Characteristics of these studies can be found in Table 10. Using a fixed-error model, the unweighted mean

TABLE 10

Characteristics of studies examining the correlation between time on homework and attitudes or conduct

Author and year	Document type	Respondent type	Sample size	Grade level	Subject matter	Correlation
Attitudes						
Cooper, Jackson, Nye, & Lindsay, 2001	Journal article	Students	214	2, 4	Student attitudes toward homework	+.03
Cooper, Lindsay, Nye, & Greathouse, 1998	Journal article	Students	709 total 285, Grades 2, 4 424, Grades 6–12	2–12	Student attitudes toward homework	+.06[a]
Hendrix, Sederberg, & Miller, 1990	Journal article	Students	1,521	12	Student attitudes concerning importance of school performance, relevance of schoolwork, effort, school-related self-esteem	+.37
Singh, Granville, & Dika, 2002	Journal article	Students	3,227	8	Student attitudes toward math and science	+.11[b]
Tonglet, 2000	Dissertation	Students	189	5, 8	Student attitudes concerning ability	+.03[c]
Conduct						
Epstein, 1988	Report	Parents	1,021	1, 3, 5	Conduct in school	−.01
Vazsonyi & Pickering, 2003	Journal article	Students	181	High school	Conduct as measured by the Normative Deviance Scale (Vasonyi & Pickering, 2000)	−.27[d]

Note. Effect sizes are coded so that positive correlations indicate that more homework is associated with more positive attitudes and fewer conduct problems.

[a]This effect size was computed from separate reported effect sizes of $r = .00$ for Grades 2 and 4 and $r = +.10$ for Grades 6–12.

[b]This effect size was computed from separate correlations between time spent on homework and how much students looked forward to math classes, $r = +.08$, and science classes, $r = +.10$, as well as the degree of usefulness that students attributed to classes in math, $r = +.10$, and science, $r = +.14$.

[c]This effect size was computed from the combined correlations between homework compliance and self-efficacy, $r = +.05$, and time spent on homework and self-efficacy, $r = +.01$.

[d]This effect size was computed from separate correlations between time spent on homework and Normative Deviance Scale scores for Caucasian students ($n = 627$), $r = +.28$, and African American students ($n = 182$), $r = +.24$.

correlation was $r = .12$. The weighted mean correlation was $r = .13$ (95% CI = .11/.14), which was significantly different from zero. Using a random-effect error model, the weighted mean correlation was $r = .13$ (95% CI = −.01/.26), not significantly different from zero.

Two studies looked at time on homework and student conduct problems. These studies are also presented in Table 10. Epstein (1988) found a near zero, $r = .01$, correlation between elementary-school parent reports of the time their child spent on homework and their conduct in school. However, Vazsonyi and Pickering (2003) found a significant negative relationship between how much time high school students reported spending on homework and their scores on the Normative Deviance Scale. Further, the relationship held for both Caucasian students, $r = .28$, and African American students, $r = .24$, separately.

Discussion

Summary of Studies on the Causal Relationship Between Homework and Achievement

Studies that have attempted to establish a causal link between homework and academic achievement have done so using several different research designs: (a) randomly assigning classrooms or students within classrooms to homework and no-homework conditions; (b) assigning homework to classrooms in a nonrandom manner but attempting statistical control of rival hypotheses; (c) using naturalistic measurement to assess both the amount of homework students do and their achievement, but attempting statistical control of rival hypotheses; and (d) testing structural equation models using naturalistic data.

The studies that randomly assigned classrooms or students within classrooms to homework and no-homework conditions were all flawed in some way that compromised their ability to draw strong causal inference. Thus we await studies that individually permit strong conclusions establishing the productive impact of homework on achievement. Still, the findings from the three studies that used random assignment did not differ in their mean effect size from the two studies that used other techniques to produce equivalent groups.

Further, the findings from manipulated-homework study designs were quite consistent and encouraging, if not conclusive. They revealed a positive relationship between homework and achievement that was robust against conservative re-analyses, including those using adjusted sample sizes and imputing possible missing data. The standardized mean difference on unit tests between students who did and did not do homework varied from $d = .39$ to $d = .97$. The weighted mean d-index was .60 under both fixed and random-error assumptions and was significantly different from zero when the student was used as the unit of analysis. When we substituted the effective sample size as the unit of analysis by adjusting for within-class dependency, the weighted mean d-index was .63 and was statistically significant, up to an assumed intraclass correlation of .35. Further, we could not reject the hypothesis that all the effect sizes from these studies were testing the same underlying population value. This was true whether fixed- or random-error assumptions were used.

Similarly, the range of estimated regression coefficients derived from studies using multiple regression, path analysis, or structural equation modeling were nearly all positive and significant. The regression coefficients appeared quite similar across subject areas. However, as with the studies described above, we would caution against drawing any conclusions regarding the mediating role of other variables on the homework–achievement relationship from this rather limited data set. The number and type of predictors in each model was complex, varied considerably from model to model, and potentially were confounded with one another across studies. Also, the estimates using naturalistic data and controlling for other variables were calculated primarily by using high school student samples.

While each set of studies is flawed, in general the studies tend not to share the same flaws. Across the set, a wide variety of students have provided data, and the effects of homework have been tested in multiple subject areas. The studies have controlled for or tested many plausible rival hypotheses in various combinations. Homework has been embedded within diverse structural models. With only rare exceptions, the relationship between the amount of homework students do and their achievement outcomes was found to be positive and statistically significant. Therefore, we think it would not be imprudent, based on the evidence in hand, to conclude that doing homework causes improved academic achievement. Of course, this assertion should not inhibit future efforts to establish more firmly this productive relationship.

The same diversity of research designs that permits optimism regarding a causal connection also makes the pinpointing of moderators of the homework–achievement relationship very problematic. Each study differs from other studies on multiple dimensions, and few studies are contained in each combination of multiple design features. This makes it difficult, if not impossible, to disentangle moderator effects by testing for plausible confounds when a moderating variable is found. Therefore, it seems unwise to use the limited data from these designs to draw inferences about what variables might be associated with the magnitude of the homework–achievement relationship. In order to get a first approximation of what these variables might be, we turn instead to an examination of a larger body of research that simply estimated the correlation between time spent on homework and achievement, without attempting to establish a causal direction for the relationship.

Summary of Homework–Achievement Correlations and Moderator Analyses

We found 69 correlations between homework and achievement reported in 32 documents. Fifty correlations were in a positive direction and 19 in a negative direction. The mean weighted correlation was $r = .24$ using a fixed-error model, and $r = .16$ using a random-error model, and both were significantly different from zero.

Moderator Analyses

It is important to keep in mind two cautions when interpreting the results of moderator analyses using correlation coefficients. First, synthesis-generated evidence should not be misinterpreted as supporting statements about causality (see Cooper, 1998). When groups of effect sizes are compared within a research synthesis, regardless of whether they come from simple correlational analyses or controlled experiments using random assignment, the synthesis can only establish an association between a moderator variable and the outcomes of studies, not a causal connection. For example, it might be found that a set of studies reporting a larger-than-average effect of homework was also conducted at upper-income schools. However, it might also be the case (known or unknown to the synthesist) that these studies tended to use unusually long homework assignments. The synthesist cannot discern which characteristic of the studies, if either, produced the larger effect. Thus, when different study characteristics are found to be associated with the effects of an intervention or the size of a correlation, the synthesist should recommend that future research examine these factors using a more systematically controlled design so that its causal impact can be appraised.

The second caution relates specifically to moderator analyses that use correlations. In the current synthesis, we are interested in the causal impact of homework on achievement. We are not interested in whether achievement also might effect time on homework (such that, for example, receiving higher grades causes students to work harder on assignments). However, we know that the size of the correlation between homework and achievement might reflect the size not only of (a) the homework-causes-achievement relationship but also of (b) the achievement-causes-homework relationship and (c) any spurious relationship between the two. Thus, unlike moderator analyses that use effect sizes from experiments, moderator analyses that use correlations must acknowledge the possibility that any uncovered relationships might be reflecting moderation of any of these three potential influences on the correlation (or that relationships involving moderators of interest are being suppressed by other relations captured by the correlation). Again, this suggests that moderator analyses in research syntheses should be interpreted with caution and used to guide future, more definitive, research.

Because of a lack of reporting or a lack of variation in some of the moderators we hoped to test, only four variables were used in quantitative analyses. Two of these, the type of outcome measure and the subject matter of the homework, revealed that time on homework was positively associated with both class grades and standardized test scores, and with reading-only, math-only, and multiple-subject outcomes. Under fixed-error assumptions, the association with homework was stronger for grades than for standardized tests, for math than for reading, and for multiple-subject outcomes than for reading and math combined. However, neither difference in association was significant under random-error assumptions, and in all instances the difference was quite small, never exceeding a difference between correlations of .06. Thus, beyond suggesting that the homework–achievement association was robust across these subsets of data, we would caution against drawing a conclusion that these moderators were important practical influences on the strength of the relation. This is especially true for subject areas because many subjects (e.g., language arts, writing, science, social studies) were not tested frequently enough to be included in the analysis.

The two other moderator variables, (a) the grade level of the student and (b) whether the student or parent reported about homework, present a different picture. For grade level, there was strong evidence that homework and achievement were positively related for secondary school students. A significant, though small, negative relationship was found for elementary school students, using fixed-error assumptions, but a nonsignificant positive relationship was found using random-error assumptions. Moreover, with both error models, the difference between the mean correlations involving elementary versus secondary students was significant.

For differences among respondents, analyses using both error models suggested that student reports about homework were significantly positively related to achievement, while parent reports produced a significant, near-zero correlation using a fixed-error model. Correlations involving the two types of respondents differed significantly. Finally, because all parent reports came from parents of elementary school students, a re-analysis of the grade-level effect was conducted excluding parent reports. This analysis still showed a higher correlation for secondary than for elementary school students under fixed-error assumptions but no difference under random-error assumptions. Not surprisingly, we also found that the correlation between time spent on homework and achievement was significantly higher when elementary school students made the report than when parents of elementary school students made the report.[4]

Explaining the Grade Level Association

There are several possible explanations for why the homework–achievement relationship differs at different grade levels. First, research in cognitive psychology indicates that age differences exist in children's ability to selectively attend to stimuli (Lane & Pearson, 1982; Plude, Enns, & Broudeur, 1994). Younger children are less able than older children to ignore irrelevant information or stimulation in their environment. Therefore, we could extrapolate that the distractions present in a young student's home environment would make home study less effective for them than for older students.

Second, younger students appear to have less effective study habits. This diminishes the amount of improvement in achievement that might be expected from homework given to them. For example, Dufresne and Kobasigawa (1989) had first-, third-, fifth-, and seventh-grade students study booklets of paired word items. They found that fifth and seventh graders spent more time studying harder items and were more likely to achieve perfect recall. Older students were also more likely to use self-testing strategies to monitor how much of the material they had learned.

At least four other explanations for the weak relationship between homework and achievement in early grades are possible. These relate more directly to the amount and purposes of homework assigned by teachers, rather than to the child's ability to benefit from study at home. Muhlenbruck, Cooper, Nye, and Lindsay (1999) found no evidence to suggest that the weaker correlation in elementary school was associated with a range restriction in the amounts of homework in early grades or that teachers assigned more homework to poorly performing classes. Evidence did suggest that teachers in early grades assigned homework more often to develop young students' management of time, a skill rarely measured on standardized achievement tests. Finally, they found some evidence that young students who were struggling in school took more time to complete homework assignments.

These last two findings suggest why the grade-level effect on homework must be viewed with caution. While it seems highly plausible to suggest that the evidence on age difference in attention span and study habits can be extrapolated to the homework situation, it is also still plausible that the relationship is due, in whole or in part, to poorer achievement in young children causing them to spend more time on homework. Or it may be that in earlier grades homework is being used for purposes other than improving immediate achievement outcomes. That is, teachers may use homework for other purposes in earlier grades because they are aware of its limited potential for improving achievement. Thus, just as we would suggest that carefully controlled studies of the causal relationship between homework and achievement be undertaken, we would also recommend that these studies include students from a variety of grade levels and that grade level be used as a moderating variable.

Explaining the Respondent Association

We can turn to some early work in social psychology, related to attribution theory, to help explain why a positive relationship between time on homework and achievement is obtained when students provide homework reports and yet the relationship hovers near zero when parents provide reports (Jones & Nisbett, 1972; Green, Lightfoot, Bandy, & Buchanan, 1985). In essence, it is likely that parents view only selected segments of their child's homework behavior. Parents are unlikely to include in their estimates of time that their children spend on homework those portions completed at school and at home before parents return from work. Parents might not even be able to accurately estimate the time that students spend on homework while both parties are home if the student does the assignments behind closed doors and alternates between homework and other activities.

We are making a case here that student reports of time on homework might be more veridical than parent reports. It is important to point out, then, that this greater veracity is based on the assumption that the *distribution* of student responses aligns better with the distribution of actual student behaviors, and not necessarily that students are not inflating their estimates. It can still be the case that students exaggerate when they report time on homework, a phenomenon that would be consistent with positive impression management. However, as long as this inflation is roughly equivalent across students (that is, students don't exchange places in the distribution), then the homework–achievement correlation can still be trusted. In essence then, our argument is that the correlation between reported and actual homework behavior is higher for student reports than parent reports. And perhaps most important, the studies that manipulated homework revealed a positive effect of homework on unit test scores. This finding is more in line with naturalistically measured student responses than parent responses. Still, it is clear that an important future direction for research would be to compare both student and parent reports about homework with behavioral observations.

Outcomes Other Than Achievement

Five studies that presented correlations between the amount of time students spent doing homework and student attitudes revealed a significant positive relationship using a fixed-error model. Two studies that looked at student conduct as an outcome produced inconsistent results. Thus, while the evidence base is small and nonexperimental in nature, it appears that the dramatic case in which large amounts of homework are cited as leading to poorer attitudes toward school and subject matter may not occur frequently enough to influence broader sample statistics.

Perhaps the most important conclusion from this synthesis regarding the effects of homework on outcomes other than achievement is that most have never been put to empirical test. While a few of the outcomes listed in Table 1 were found in the studies covered herein and some others can be found in research that examines homework from different perspectives (e.g., Hoover-Dempsey et al., 2001), the majority of these outcomes remain fertile ground for future research.

A Comparison With the Results of Cooper (1989)

Cooper (1989) reported an average effect size of $d = .21$ from studies that compared students who did homework with students who did no homework. This synthesis found a mean effect size of $d = .60$. These results suggest much larger effects in more recent studies. Looking for potential sources of the difference suggests that the research designs and achievement measures across the two syntheses might not be directly comparable. For example, the 16 studies listed in Cooper's (1989) Table 5.3 (p. 66) included four studies that used nonequivalent control groups without matching. All of the studies in the current synthesis used some form of matching. This might have improved their ability to detect a homework effect. Also, four of the studies in the earlier synthesis used standardized tests as the outcome measure. In the current synthesis, all of the studies used unit tests, a measure more closely aligned with the content of the homework assignments. Regardless, it seems clear that more recent studies that introduced homework as an exogenous intervention have revealed more impressive effects of homework.

Cooper (1989) reported a mean effect size of $d = .09$ for studies that compared homework with in-school supervised study. We found no study conducted since 1987 that carried out a similar type of comparison. Conversely, we uncovered numerous naturalistic studies that controlled for other variables confounded with the homework–achievement relationship and found these to reveal near-uniform positive results. Cooper (1989) did not include this type of research design in the earlier synthesis.

Finally, the Cooper (1989) synthesis reported a mean simple correlation of $r = .19$ between homework and achievement using a fixed-error model. We found the corresponding correlation to be $r = .24$. Thus these estimates appear very consistent.

Optimum Amounts of Homework

Cooper's (1989) meta-analysis found that for high school students the positive relation between time on homework and achievement did not appear until at least 1 hour of homework per week was reported. Then the linear relation continued to climb unabated to the highest measured interval (more than 2 hours per night). For junior high students the positive relation appeared for even small amounts of time on homework (less than 1 hour per night) but disappeared entirely after students reported doing between 1 and 2 hours each night. Only one study was available for Grades 1–6 but the lack of a simple linear relationship at these grade levels suggested the line would be flat.

We found only one study that permitted interpretation regarding optimum homework amounts. Lam (1996) found that for Caucasian American and Asian American high school students the strongest relationship between homework and achievement was found among students reporting doing 7 to 12 hours of homework per week, followed by students reporting doing 13–20 hours per week. This finding extends the conclusions from the earlier synthesis because it was not able to make a distinction in time spent on homework per night beyond 2 hours for high school students. Assuming that the causal direction of these findings is predominantly one in which more homework causes better achievement, the Lam (1996) finding suggests that the optimum benefits of homework for high school students might lie between 1½ and 2½ hours. Of course, there is still much guesswork in these estimates, and optimum amounts of homework likely will be dependent on many factors, including the nature of the assignment and student individual differences. Also, the Lam (1996) study was limited to 12th-grade Chinese Americans and Caucasian Americans. Still, this new piece of evidence does suggest, as common sense would dictate, that the positive effects of homework are not linear across all amounts. Even for these oldest students, too much homework may diminish its effectiveness, or even become counterproductive.

Limitations of Generalizability

Our analyses looking for publication bias and data censoring revealed little evidence to suggest that the strategies we used to locate studies for the synthesis were in some way a biased representation of all studies that might exist. That being said, it is also the case that certain clear limitations to the generalizability of the synthesis findings need to be noted.

First, as noted above, the positive causal effect of homework on achievement has been tested and found only on measures of an immediate outcome, the unit test. Therefore, it is not possible to make claims about homework's causal effects on longer-term measures of achievement, such as class grades and standardized tests, or other achievement-related outcomes. However, the studies using naturally occurring measures of time on homework found strong evidence of a link to longer-term achievement measures. We suspect that this distinction in the types of measures used in experimental and naturalistic studies of homework will persist. This is because the large-scale manipulation of homework across multiple subject areas and long durations within the same samples of students—the type of experiment likely needed to produce homework effects on grades and standardized tests—will require considerable resources and the cooperation of educators and parents willing to participate.

With regard to subject matter, both studies that introduced homework as an exogenous intervention and studies that used statistical controls suggest that homework will have positive effects on achievement involving both quantitative and verbal material. However, our database contained too few correlations involving other subjects, such as science and social studies, to include them in the meta-analysis. Therefore, while there is evidence that the effect of subject matter on the homework–achievement relationship is small, it should be viewed as suggestive rather than conclusive.

Finally, a perusal of Tables 3 through 8 suggests that few studies exist examining the effectiveness of homework in the early elementary school grades. This may be an especially important omission because of the apparent increase in the amount of homework being assigned to students in these grades (Hofferth & Sandberg, 2000). Also, nearly all the literature that we uncovered looked at the effect of homework on students who might be labeled "average," or examined broad samples of students but did not look for moderating effects of student characteristics.

Future Research

Throughout this discussion, we have pointed to fruitful avenues for future research. As is often the case, an assessment of what we know places in bold relief what we don't. Researchers are encouraged to find in our report any of the numerous areas where research is thin or nonexistent. These areas include studies that introduce homework as an exogenous intervention, randomly assign students or classrooms to conditions, and then analyze data at the same unit of analysis as the manipulation. There are several barriers to implementing such designs. First, of course, are the barriers to random assignment in applied settings (see Shadish, Cook, & Campbell, 2002, pp. 287–288), not the least of which would be the ethics of withholding from some students an intervention (homework) with presumed benefits. Second, if treatments are implemented at the classroom level and analyzed accordingly, the statistical power to detect effects will be quite low unless large-scale studies can be mounted that involve numerous classrooms. If students within classrooms are assigned to conditions, the researcher faces issues of treatment diffusion and/or demoralization and compensation effects that can contaminate conditions, because the intervention and control groups interact and know each other's experimental assignment.

Still, given the state of evidence, it seems there is much less to be gained from carrying out "homework studies as usual" than from new attempts to pinpoint estimates of causal relationships. That being said, we would encourage, as well, the use of mixed research models that incorporate qualitative analyses—to examine the homework process, moderators, and mediators of its effects, along with its intended *and* unintended consequences—in experimental designs. Such studies provide a rich tableau and complementary sources of knowledge for guiding yet another generation of research, policy, and practice. The long-term and cumulative effects of homework remain a largely unmapped terrain. Therefore, nonexperimental, longitudinal studies that follow cohorts of students and perform fine-grained analyses of developing homework behaviors would be a new and rich source of information.

In addition, the gaps in our knowledge suggest that future studies, whether experimental, qualitative, or longitudinal, should include variations in numerous potential factors in homework effects. Most important, we think these variations should include:

1. Students in multiple grades, especially the early elementary grades;
2. Students with other varying characteristics, especially varying ability levels, SES, and sex;
3. Variations in the subject matter of homework assignments, including subjects other than reading and math;
4. Measures of the non-achievement-related effects of homework that have been proposed in the literature; and
5. Variations in the amount of homework assigned, so that optimum amounts of homework can be examined.

We might envision all of these design variations being realized within a single research project, leading to multiple replications, but it is more likely that numerous small projects will gather data on one or a few areas. Thus we more realistically call for *programs of research* that begin by establishing general principles (some of which can be gleaned from this synthesis) and then systematically vary factors 1 through 5, above, not in the same study but through a series of interrelated studies (Shadish et al., 2002). The advantages of this approach are that studies can be implemented with good control of treatment, thus enhancing their power to detect effects. And, of course, individual studies will be less expensive to conduct. A disadvantage of this approach is that it limits the ability to examine interactions between factors. For example, if the grade level of the student is examined in one small study and sex of the student in another, it is impossible to examine whether the students' grade level and sex interact in their moderation of homework's effects.

Conclusion

We hope that this report has demonstrated the value of research synthesis for testing the plausibility of causal relationships even when less-than-optimal research designs and analyses are available in the literature. Most important, we hope that the findings provide the beginnings of an empirical foundation on which educators can base homework policies and practices and researchers can build the next generation of homework research.

Notes

This research was supported by a grant from the U.S. Department of Education, Office of Educational Research and Improvement (R305T030002), to the first author. The opinions expressed herein are those of the authors and not necessarily of the Department of Education. The authors wish to thank Heather Meggers-Wright and Jeffrey C. Valentine for their assistance with conducting the research synthesis, and Angela Clinton for assistance with manuscript preparation. Correspondence regarding this article should be sent to Harris Cooper (see contact information at end of article).

[1]We could find no study that looked at the students' SES as a moderator of the homework–achievement link. Only two studies examined the sex of the student as a moderator of the homework–achievement link. Among the studies that manipulated homework, Foyle (1984) presented results of an Analysis of Covariance that included the sex of the student in interaction with homework condition. The interaction was not significant, and the cross-break of means was not reported. Among the studies reporting simple correlations, Mau and Lynn (2000) reported six comparisons of male and female correlations between homework and achievement in Grades 10 and 12 for math, reading, and science. All six comparisons revealed significantly higher correlations for females than for males.

[2]Looking for missing correlations to the right (increasing the size of the positive effect) suggested more evidence that correlations higher than those in the retrieved reports might have been missing from the data set. The fixed-error model suggested that 11 correlations might be missing and that if they were imputed, fixed graph $r = .25$ (95% CI = .25/.26). The random model imputed no additional correlations. Also, the trim-and-fill analysis was conducted separately for studies that used class grades or standardized achievement tests as outcome variables. In all cases, the analysis suggested that the findings reported in this article were robust with regard to data censoring.

[3]The Cooper et al. (1998) correlation involving parents had to be dropped from this analysis because it included both elementary and secondary students.

[4]This type of subgroup analysis is a way of disentangling the effects of confounded moderating variables. It is an example of the type of analysis that would have been beneficial to carry out as well on the studies that employed exogenous introductions of homework. However, the limited number of such studies meant that some combinations of categories within moderators would have few or no studies in them.

References

References marked with an asterisk indicate studies included in the meta-analysis.

Acock, A. C., & Demo, D. H. (1994). *Family diversity and well being*. Thousand Oaks, CA: Sage

Alleman, J., & Brophy, E. J. (1991). *Reconceptualizing homework as out-of-the-school learning opportunities* (Occassional Paper No. 135). East Lansing, MI: Michigan State University, Institute for Research in Teaching.

*Antonek, J. L. (1996). Interactive homework: Linking parents to the foreign language learning of their children (Doctoral dissertation, University of Pittsburgh, 1996). *Dissertation Abstracts International, 57,* 5084.

Balli, S. J. (1998). When Mom and Dad help: Student reflections on parent involvement with homework. *Journal of Research and Development in Education, 31,* 142–146.

Balli, S. J., Demo, D. H., & Wedman, J. F. (1998). Family involvement with children's homework: An intervention in the middle grades. *Family Relations, 47,* 149–157.

Balli, S. J., Wedman, J. F., & Demo, D. H. (1997). Family involvement with middle-grades homework: Effects of differential prompting. *Journal of Experimental Education, 66,* 31–48.

Barnett, V., & Lewis, T. (1994). *Outliers in statistical data analysis* (3rd ed.). New York: John Wiley & Sons.

Baumgartner, D., Bryan, T., Donahue, M., & Nelson, C. (1993). Thanks for asking: Parent comments about homework, tests, and grades. *Exceptionality: A Research Journal, 4,* 177–185.

Becker, H. J., & Epstein, J. L. (1982). Parent involvement: A survey of teacher practices. *Elementary School Journal, 83,* 85–102.

*Bents-Hill, C., Boswell, R., Byers, J., Cohen, N., Cummings, J., & Leavitt, B. (1988, April). *Relationship of academic performance to parent estimate of homework time.* Paper presented at the annual meeting of the National Association of School Psychologists, Chicago.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). Comprehensive Meta Analysis (Version 2.1) [Computer software]. Englewood, NJ: BioStat.

*Bowen, N. K., & Bowen, G. L. (1998). The mediating role of educational meaning in the relationship between home academic culture and academic performance. *Family Relations, 47,* 45–51.

*Brookhart, S. M. (1997). Effects of classroom assessment environment on mathematics and science achievement. *Journal of Educational Research, 90,* 323–330.

*Broxie, G. J. B. (1987). Parental participation in homework completion as a predictor of academic success of students in the single parent family (Doctoral dissertation, Virginia Polytechnic Institute and State University, 1987). *Dissertation Abstracts International, 51,* 2806.

*Bruce, F. A., Jr., & Singh, K. (1996). Academic achievement: A model of school learning for eighth grade students. *Research in Middle Level Education Quarterly, 19,* 95–111.

Bryan, T., Nelson, C., & Mathru, S. (1995). Homework: A survey of primary students in regular, resource, and self-contained special education classrooms. *Journal of Learning Disabilities, 27,* 85–90.

*Camp, W. G. (1990). Participation in student activities and achievement: A covariate structural analysis. *Journal of Educational Research, 83,* 272–278.

Campbell, J. R., Reese, C. M., O'Sullivan, C., & Dossey, J. A. (1996). *NAEP 1994 Trends in Academic Progress.* Washington, DC: U.S. Department of Education.

Carroll, J. (1963). A model of school learning. *Teachers College Record, 64,* 723–733.

Chen, C., & Stevenson, H. W. (1989). Homework: A cross-cultural examination. *Child Development, 60,* 551–561.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Lawrence Erlbaum.

Coleman, J. S., Hoffer, T., & Kilgore, S. (1982). *High school achievement.* New York: Basic Books.

*Cool, V. A., & Keith, T. Z. (1991). Testing a model of school learning: Direct and indirect effects on academic achievement. *Contemporary Educational Psychology, 16,* 28–44.

Cooper, H. (1989). *Homework.* White Plains, NY: Longman.

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

Cooper, H., & Hedges, L. B. (Eds.). (1994). *Handbook of research synthesis.* New York: Russell Sage Foundation.

*Cooper, H., Jackson, K., Nye, B., & Lindsay, J. J. (2001). A model of homework's influence on the performance of elementary school students. *Journal of Experimental Education, 69,* 181–199.

*Cooper, H., Lindsay, J. J., Nye, B., & Greathouse, S. (1998). Relationships among attitudes about homework, amount of homework assigned and completed, and student achievement. *Journal of Educational Psychology, 90,* 70–83.

Cooper, H., Lindsay, J. J., & Nye, B. (2000). Homework in the home: How student, family, and parenting-style differences relate to the homework process. *Contemporary Educational Psychology, 25,* 464–487.

Corno, L. (1994). Student volition and education: Outcomes, influences, and practices. In D. H. Schunk & B. J. Zimmerman (Eds.), *Self-regulation of learning and performance: Issues and educational applications* (pp. 229–251). Hillsdale, NJ: Lawrence Erlbaum.

Corno, L. (1996). Homework is a complicated thing. *Educational Researcher, 25*(8), 27–30.

Corno, L., & Xu, J. (1998, April). *Homework and personal responsibility.* Paper presented at the annual meeting of the American Educational Research Association, San Diego.

Coutts, P. M. (2004). Meanings of homework and implications for practice. *Theory into practice, 43,* 182–188.

Covington, M. V. (1998). *The will to learn: A guide for motivating young people.* Cambridge, UK: Cambridge University Press.

*Davis, J. E., & Jordan, W. J. (1994). The effects of school context, structure, and experiences on African American males in middle and high school. *Journal of Negro Education, 63,* 570–587.

Deslandes, R., & Cloutier, R. (2002). Adolescents' perception of parental involvement in schooling. *School Psychology International, 23*(2), 220–232.

*Deslandes, R., Royer, E., Potvin, P., & Leclerc, D. (1999). Patterns of home and school partnership for general and special education students at the secondary level. *Exceptional Children, 65,* 496–506.

*Drazen, S. (1992). *Student achievement and family and community poverty: Twenty years of education reform.* Paper presented at the American Educational Research Association, Boston.

Dufresne, A., & Kobasigawa, A. (1989). Children's spontaneous allocation of study time: Differential and sufficient aspects. *Journal of Experimental Child Psychology, 42,* 274–296.

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

Duval, S., & Tweedie, R. (2000b). Trim and fill: A simple funnel plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56,* 276–284.

Epps, M. (1966). *Homework.* Washington, DC: National Education Association.

*Epstein, J. L. (1988). *Homework practices, achievements, and behaviors of elementary school students.* Baltimore: Center for Research on Elementary and Middle Schools. (ERIC Document Reproduction Service No. ED301322)

Epstein, J. L. (2001). *School, family, and community partnerships: Preparing educators and improving schools.* Boulder, CO: Westview.

Epstein, J. L., & Dauber, S. L. (1991). School programs and teacher practices of parent involvement in inner-city elementary and middle schools. *Elementary School Journal, 91,* 289–305.

Epstein, J. L., Simon, B. S., & Salinas, K. C. (1997). *Effects of Teachers Involve Parents in Schoolwork (TIPS) language arts interactive homework in the middle grades.* Research Bulletin 18 (September). Bloomington, IN: Phi Delta Kappa, CEDR.

Epstein, J. L., & Van Voorhis, F. L. (2001). More than minutes: Teachers' roles in designing homework. *Educational Psychologist, 36,* 181–193.

Farkas, S., Johnson, J. M., & Duffet, A. (1999). *Playing their parts: Parents and teachers talk about parental involvement in public schools.* New York: Public Agenda.

*Fehrman, P. G., Keith, T. Z., & Reimers, T. Z. (1987). Home influence on school learning: Direct and indirect effects of parental involvement on high school grades. *Journal of Educational Research, 80,* 330–337.

*Finstad, E. (1987). *Effects of mathematics homework on second grade achievement.* Houston, TX: Sam Houston State University. (ERIC Document Reproduction Service No. ED291609)

*Foyle, H. A. (1984). The effects of preparation and practice homework on student achievement in tenth-grade American history (Doctoral dissertation, Kansas State University, 1984). *Dissertation Abstracts International, 45,* 2474.

*Foyle, H. (1990). *Homework and cooperative learning: A classroom field experiment.* Emporia, KS: Emporia State University, Faculty Research and Creativity Committee. (ERIC Document No. ED350285)

Gill, B., & Schlossman, S. (1996). "A sin against childhood": Progressive education and the crusade to abolish homework, 1897–1941. *American Journal of Education, 105,* 27–66.

Gill, B. P., & Schlossman, S. L. (2000). The lost cause of homework reform. *American Journal of Education, 109,* 27–62.

Gill, B., & Schlossman, S. (2003). A nation at rest: The American way of homework. *Educational Evaluation and Policy Analysis, 25,* 319–337.

Goldstein, A. (1960). Does homework help? A review of research. *Elementary School Journal, 60,* 212–224.

González, N., Andrade, R., Civil, M., & Moll, L. (2001). Bridging funds of distributed knowledge: Creating zones of practice in mathematics. *Journal of Education of Students Placed at Risk, 6,* 115–132.

Green, S. K., Lightfoot, M. A., Bandy, C., & Buchanan, D. R. (1985). A general model of the attribution process. *Basic and Applied Social Psychology, 6,* 159–179.

Greenhouse, J. B., & Iyengar, S. (1994). Sensitivity analysis and diagnostics. In H. Cooper & L. V. Hedges (Eds.), *Handbook of research synthesis* (pp. 383–398). New York: Russell Sage Foundation.

Grubbs, F. E. (1950). Sample criteria for testing outlying observations. *Journal of the American Statistical Association, 21,* 27–58.

Hagan, H. (1927). The value of homework as compared with supervised study. *Second Yearbook, Chicago Principals' Club,* 147–149. Chicago: Chicago Principals' Club.

Harris, S., Nixon, J., & Rudduck, J. (1993). School work, homework and gender. *Gender and Education, 5*(1), 3–14.

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

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

Henderson, M. (1996). *Helping your student get the most out of homework.* Washington, DC: National Education Association.

*Hendrix, V. L., Sederberg, C. H., & Miller, V. L. (1990). Correlates of commitment/alienation among high school seniors. *Journal of Research and Development in Education, 23,* 129–135.

*Hightower, L. B. (1991). The influence of part time work and viewing television on homework: The delicate network of variables that affect student achievement (Doctoral dissertation, University of Arkansas, 1991). *Dissertation Abstracts International, 53,* 2758.

*Hill, A. T. (2003). A panel study of the effects of teacher education, class size, and time-on-task on student achievement (Doctoral dissertation, University of Delaware, 2003). *Dissertation Abstracts International, 64,* 1017.

Hofferth, S. L., & Sandberg, J. F. (2000). *Changes in American children's time, 1981–1997.* Ann Arbor: University of Michigan Population Studies Center.

Hong, E., & Milgram, R. M. (1999). Preferred and actual homework style: A cross-cultural examination. *Educational Research, 41*(3), 251–265.

Hoover-Dempsey, K. V., Bassler, O. C, & Burow, R. (1995). Parents' reported involvement in students' homework: Parameters of reported strategy and practice. *Elementary School Journal, 95,* 435–450.

Hoover-Dempsey, K. V., Battiato, A. C., Walker, J. M. T., Reed, R. P., DeJong, J. M., & Jones, K. P. (2001). Parental involvement in homework. *Educational Psychologist, 36,* 195–209.

Jackson, C. (2003). Motives for "laddishness" at school: Fear of failure and fear of the "feminine." *British Educational Research Journal, 29*(4), 583–598.

Johnson, J. K., & Pontius, A. (1989). Homework: A survey of teacher beliefs and practices. *Research in Education, 41,* 71–78.

Jones, E. E., & Nisbett, R. E. (1972). The actor and observer: Divergent perceptions of the causes of behavior. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 79–94). Morristown, NJ: General Learning.

Jones, R. D., & Colvin, R. (1964). Abolish homework: Let supervised school work take its place. *Clearing House, 39,* 206–209.

*Keith, T. Z., & Benson, M. J. (1992). Effects of manipulable influences on high school grades across five ethnic groups. *Journal of Educational Research, 86,* 85–93.

*Keith, T. Z., & Cool, V. A. (1992). Testing models of school learning: Effects of quality of instruction, motivation, academic coursework, and homework on academic achievement. *School Psychology Quarterly, 7,* 207–226.

*Keith, T. Z., Diamond-Hallam, C., & Fine, J. G. (2004). Longitudinal effects of in-school and out-of-school homework on high school grades. *School Psychology Quarterly, 19,* 187–211.

*Keith, T. Z., & Keith, P. B. (1993). Does parental involvement affect eighth-grade student achievement? Structural analysis of national data. *School Psychology Review, 22,* 474–495.

*Keith, T. Z., Keith, P. B., Bickley, P. G., & Singh, K. (1992). *Effects of parental involvement on eighth grade achievement: LISREL analysis of NELS-88 data.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Kouzma, N. M., & Kennedy, G. A. (2002). Homework, stress, and mood disturbance in senior high school students. *Psychological Reports, 91,* 193–198.

Kralovec, E., & Buell, J. (2000). *The end of homework: How homework disrupts families, overburdens children, and limits learning.* Boston: Beacon Press.

*Lam, J. W. (1996). The employment activity of Chinese-American high school students and its relationship to academic achievement (Master's thesis, University of Texas at Arlington, 1996). *Masters Abstracts International, 34,* 2148.

Lane, D. M., & Pearson, D. A. (1982). The development of selective attention. *Merrill-Palmer Quarterly, 28,* 317–337.

Lee, J. F., & Pruitt, K. W. (1979). Homework assignments: Class games or teaching tools? *Clearing House, 53,* 31–35.

Levin, I., Levy-Shiff, R., Appelbaum-Peled, T., Katz, I., Komar, M., & Meiran, N. (1997). Antecedents and consequences of maternal involvement in children's homework: A longitudinal analysis. *Journal of Applied Developmental Psychology, 18,* 207–227.

Lindsay, C. (1928). Home study versus class room instruction. *Mississippi Educational Advance, 20,* 88–89.

Loveless, T. (2003). *The 2003* Brown Center *Report on American* education: *How well are American students learning? With sections on homework, charter schools, and rural school achievement.* Brookings Institution Press: Washington, DC.

*Mau, W. C., & Lynn, R. (2000). Gender differences in homework and test scores in mathematics, reading and science at tenth and twelfth grade. *Psychology, Evolution & Gender, 2,* 119–125.

McDermott, R. P., Goldman, S. V., & Varenne, H. (1984). When school goes home: Some problems in the organization of homework. *Teachers College Record, 85,* 391–409.

*McGrath, J. B. (1992). Student and parental homework practices and the effect of English homework on student test scores. *Dissertation Abstracts International, 53* (10A), 3490. (UMI No. 9231359)

*Meloy, L. L. (1987). Effects of homework on language arts achievement in third and fourth grades (Doctoral dissertation, University of Iowa, 1987). *Dissertation Abstracts International, 51,* 725.

*Modi, M., Konstantopoulos, S., & Hedges, L. V. (1998). *Predictors of academic giftedness among U.S. high school students.* Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Mulhenbruck, L., Cooper, H., Nye, B., & Lindsay, J. J. (1999). Homework and achievement: Explaining the different strengths of relation at the elementary and secondary school levels. *Social Psychology of Education, 3,* 295–317.

Nash, J. B. (1930). What price home study? *The School Parent, 9,* 6–12.

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform.* Washington, DC: U.S. Department of Education.

Odum, M. E. (1994, May 2). A tough assignment for teachers: Getting students to do homework is getting harder. *Washington Post,* A1, A10.

*Olson, M. (1988). Homework variables and academic achievement: An integrated study. *Dissertation Abstracts International, 49,* 08A. (UMI No. 8821110)

Otto, H. J. (1941). Elementary education. In W. S. Monroe (Ed.), *Encyclopedia of education research* (pp. 444–445). New York: Macmillan.

Paschal, R., Weinstein, T., & Walberg, H. (1984). The effects of homework on learning: A quantitative synthesis. *Journal of Educational Research, 78,* 97–104.

Patall, E. A., Cooper, H., & Robinson, J. C. (2005). Parent involvement in homework: A synthesis of research. Manuscript in preparation.

Patri, A. (1925). *School and home.* New York: Appleton.

Paula, N. (1995). *Helping your child with homework.* Washington, DC: Office of Educational Research and Improvement, U.S. Department of Education.

*Peng, S. S., & Wright, D. (1994). Explanation of academic achievement of Asian American students. *Journal of Educational Research, 87,* 346–352.

*Pezdek, K., Berry, T., & Renno, P. A. (2002). Children's mathematics achievement: The role of parents' perceptions and their involvement in homework. *Journal of Educational Psychology, 94,* 771–777.

Plude, D. J., Enns, J. T., & Brodeur, D. (1994). The development of selective attention: A life-span overview. *Acta Psychologica, 86,* 227–272.

*Portes, A., & MacLeod, D. (1996). Educational progress of children of immigrants: The roles of class, ethnicity, and school context. *Sociology of Education, 69,* 255–275.

*Portes, P. R., & Zady, M. F. (2001). *Determinants of reading achievement of immigrant adolescents: The role of demographic and psycho-cultural factors in school adaptation.* Paper presented at the annual meeting of the American Educational Research Association, Seattle.

Public Agenda. (2000). *Survey finds little sign of backlash against academic standards or standardized tests.* New York: Author.

Ratnesar, R. (1999, January 25). The homework ate my family. *Time, 153,* 54–63.

Reese, W. (1995). *The origins of the American high school.* New Haven, CT: Yale University Press.

*Roberts, J. K. (2000). Examining mutable reform options for urban schools with multilevel analysis in the National Education Longitudinal Study: 88 (Doctoral dissertation, Texas A & M University, 2000). *Dissertation Abstracts International, 61,* 2597.

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *Handbook of research synthesis* (pp. 231–244). New York: Russell Sage.

*Rozevink, M. L. (1995). Selected small school motivational factors regarding homework completion in grades nine and twelve (Doctoral dissertation, Bowling Green State University, 1995). *Dissertation Abstracts International, 57,* 59.

Rutter, M., Maughan, B., Mortimore, P., & Ouston, J. (1979). *Fifteen thousand hours: Secondary schools and their effects on children.* Cambridge, MA: Harvard University Press.

San Diego City Schools Research Department. (1936). *Home study practices and opinions.* San Diego, CA: San Diego Board of Education.

*Schewior, R. W. (2001). Educational outcomes and student effort: Evidence from NELS 1992 (Doctoral dissertation, Rutgers University, 2001). *Dissertation Abstracts International, 62* (09A), 3140. (UMI No. 3026339)

Scott-Jones, D. (1984). Family influences on cognitive development and school achievement. *Review of Research in Education, 11,* 259–304.

Scott-Jones, D. (1995). Parent-child interactions and school achievement. In B. A. Ryan, G. R. Adams, T. P. Gullota, R. P. Weissberg, & R. L. Hampton (Eds.), *The family–school connection: Theory, research, and practice* (pp. 75–107). Thousand Oaks, CA: Sage.

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

Shadish, W. R., & Haddock, C. K. (1994). Combining estimates of effect size. In H. Cooper & L. V. Hedges (Eds.), *Handbook of research synthesis* (pp. 383–398). New York: Russell Sage Foundation.

*Singh, K., Granville, M., & Dika, S. (2002). Mathematics and science achievement: Effects of motivation, interest, and academic engagement. *Journal of Educational Research, 95,* 323–332.

*Smith, T. E. (1990). Time and academic achievement. *Journal of Youth and Adolescence, 19,* 539–558.

*Smith, T. E. (1992). Time use and change in academic achievement: A longitudinal follow-up. *Journal of Youth and Adolescence, 21,* 725–747.

Thayer, V. (1928). *The passing of the recitation.* New York: Heath.

*Thomas, J. P. (2001). *A model of mathematics achievement using proficiency scores.* Graylake, IL: College of Lake County. (ERIC Document Reproduction Service No. ED459216)

*Thomas, J. P. (2002). *An examination of the relationship between learning variables and academic achievement outcomes.* (ERIC Documentation Reproduction Service No. ED471988)

*Tonglet, J. P. (2000). Influences on math homework completion and achievement: Students' attitudes toward teacher-related factors, student motivational factors, and environment-related factors in fifth and eighth graders (Doctoral dissertation, Tulane University, 2000). *Dissertation Abstracts International, 61,* 3468.

*Townsend, S. (1995). The effects of vocabulary homework on third grade achievement (Master's thesis, Kean College of New Jersey, 1995). (ERIC No. ED379643).

Van Voorhis, F. (2003). Interactive homework in middle school: Effects on family involvement and science achievement. *Journal of Educational Research, 96,* 323–338.

*Vazsonyi, A. T., & Pickering, L. E. (2003). The importance of family and school domains in adolescent deviance: African American and Caucasian youth. *Journal of Youth and Adolescence, 32,* 115–128.

Walberg, H., & Paschal, R. (1995). Homework. In L. Anderson (Ed.), *International Encyclopedia of Teaching and Teacher Education* (pp. 268–271). Oxford, UK: Elsevier.

*Walker, D. R. (2002). African American males: What do schools and teachers do to enhance academic achievement (Doctoral dissertation, University of North Carolina at Charlotte, 2002). *Dissertation Abstracts International, 63,* 88.

Warton, P. M. (1998). Mothers' views about homework in the early years of school. *Australian Journal of Early Childhood, 23,* 35–39.

Warton, P. M. (2001). The forgotten voice in homework: Views of Students. *Educational Psychologist, 36,* 155–165.

Winerip, M. (1999, January 3). Too much too soon or too little too late? A report from the homework front. *New York Times, Education Life,* 28–31, 42.

*Wynn, J. L. (1996). The effects of family involvement, homework, and TV time on student grade point average (Doctoral dissertation, Auburn University, 1996). *Dissertation Abstracts International, 57,* 3356.

*Wynstra, J. E. (1995). The relationship between five aspects of the home environment and students reading above grade level (Master's thesis, Cumberland College, 1995). (ERIC Document Reproduction No. CS012891)

Xu, J., & Corno, L. (1998). Case studies of families doing third grade homework. *Teachers College Record, 100,* 402–436.

Younger, M., & Warrington, M. (1996). Differential achievement of girls and boys at GCSE: Some observations from the perspective of one school. *British Journal of Sociology of Education, 17*(3), 299–313.

Zimmerman, B., Bonner, S., & Kovach, R. (1996). *Developing self-regulated learners: Beyond achievement to self-efficacy.* Washington, DC: American Psychological Association.

Authors

HARRIS COOPER is a Professor of Psychology and Director of the Program in Education, Box 90739, Duke University, Durham, NC 27708-0739; e-mail *cooperh@duke.edu.* His research interests include how academic activities outside the school day (such as homework, after school programs, and summer school) affect the achievement of children and adolescents; he also studies techniques for improving research synthesis methodology and how to make research synthesis more useful for practitioners and policymakers.

JORGIANNE CIVEY ROBINSON is a PhD candidate in Social Psychology in the Department of Psychology, Box 90085, Duke University, Durham, NC 27708-0085; e-mail *jorgianne.robinson@duke.edu.* Her research interests include homework and the impact of variations in assignments on homework effectiveness.

ERIKA A. PATALL is a graduate student in Social Psychology in the Department of Psychology, Box 90085, Duke University, Durham, NC 27708-0085; e-mail *erika.patall@duke.edu.* She is also a fellow in the Spencer Foundation Interdisciplinary Program in Education Science. Her research interests include parent involvement in homework and the role of choice in intrinsic motivation.

Exhibit 1.9

(page intentionally left blank)

The Case for (Quality) Homework

Why it improves learning, and how parents can help

Janine Bempechat

PHOTOGRAPH / PACO NAVARRO; GETTY IMAGES

Any parent who has battled with a child over homework night after night has to wonder: Do those math worksheets and book reports really make a difference to a student's long-term success? Or is homework just a headache—another distraction from family time and downtime, already diminished by the likes of music and dance lessons, sports practices, and part-time jobs?

Allison, a mother of two middle-school girls from an affluent Boston suburb, describes a frenetic afterschool scenario: "My girls do gymnastics a few days a week, so homework happens for my 6th grader after gymnastics, at 6:30 p.m. She doesn't get to bed until 9. My 8th grader does her homework immediately after school, up until gymnastics. She eats dinner at 9:15 and then goes to bed, unless there is more homework to do, in which case she'll get to bed around 10." The girls miss out on sleep, and weeknight family dinners are tough to swing.

Parental concerns about their children's homework loads are nothing new. Debates over the merits of homework—tasks that teachers ask students to complete during non-instructional time—have ebbed and flowed since the late 19th century, and today its value is again being scrutinized and weighed against possible negative impacts on family life and children's well-being.

Are American students overburdened with homework? In some middle-class and affluent communities, where pressure on students to achieve can be fierce, yes. But in families of limited means, it's often another story. Many low-income parents value homework as an important connection to the school and the curriculum—even as their children report receiving little homework. Overall, high-school students relate that they spend less than one hour per day on homework, on average, and only 42 percent say they do it five days per week. In one recent survey by the National Assessment of Educational Progress (NAEP), a minimal 13 percent of 17-year-olds said they had devoted more than two hours to homework the previous evening (see Figure 1).

Recent years have seen an increase in the amount of homework assigned to students in grades K–2, and critics point to research findings that, at the elementary-school level, homework does not appear to enhance children's learning. Why, then, should we burden young children and their families with homework if there is no academic benefit to doing it? Indeed, perhaps it would be best, as some propose, to eliminate homework altogether, particularly in these early grades.

On the contrary, developmentally appropriate homework plays a critical role in the formation of positive learning beliefs and behaviors, including a belief in one's academic ability, a deliberative and effortful approach to mastery, and higher expectations and aspirations for one's future. It can prepare children to confront ever-more-complex tasks, develop resilience in the face of difficulty, and learn to embrace rather than shy away from challenge. In short, homework is a key vehicle through which we can help shape children into mature learners.

The Homework-Achievement Connection

A narrow focus on whether or not homework boosts grades and test scores in the short run thus ignores a broader purpose in education, the development of lifelong, confident learners. Still, the question looms: *does* homework enhance academic success? As the educational psychologist Lyn Corno wrote more than two decades ago, “homework is a complicated thing.” Most research on the homework-achievement connection is correlational, which precludes a definitive judgment on its academic benefits. Researchers rely on correlational research in this area of study given the difficulties of randomly assigning students to homework/no-homework conditions. While correlation does not imply causality, extensive research has established that at the middle- and high-school levels, homework completion is strongly and positively associated with high achievement. Very few studies have reported a negative correlation.

As noted above, findings on the homework-achievement connection at the elementary level are mixed. A small number of experimental studies have demonstrated that elementary-school students who receive homework achieve at higher levels than those who do not. These findings suggest a causal relationship, but they are limited in scope. Within the body of correlational research, some studies report a positive homework-achievement connection, some a negative relationship, and yet others show no relationship at all. Why the mixed findings? Researchers point to a number of possible factors, such as developmental issues related to how young children learn, different goals that teachers have for younger as compared to older students, and how researchers define homework.

Certainly, young children are still developing skills that enable them to focus on the material at hand and study efficiently. Teachers’ goals for their students are also quite different in elementary school as compared to secondary school. While teachers at both levels note the value of homework for reinforcing classroom content, those in the earlier grades are more likely to assign homework mainly to foster skills such as responsibility, perseverance, and the ability to manage distractions.

Most research examines homework generally. Might a focus on homework in a specific subject shed more light on the homework-achievement connection? A recent meta-analysis did just this by examining the relationship between math/science homework and achievement. Contrary to previous findings, researchers reported a stronger relationship between homework and achievement in the elementary grades than in middle school. As the study authors note, one explanation for this finding could be that in elementary school, teachers tend to assign more homework in math than in other subjects, while at the same time assigning shorter math tasks more frequently. In addition, the authors point out that parents tend to be more involved in younger children’s math homework and more skilled in elementary-level than middle-school math.

In sum, the relationship between homework and academic achievement in the elementary-school years is not yet established, but eliminating homework at this level would do children and their families a huge disservice: we know that children’s learning beliefs have a powerful impact on their academic outcomes, and that through homework, parents and teachers can have a profound influence on the development of positive beliefs.

How Much Is Appropriate?

Harris M. Cooper of Duke University, the leading researcher on homework, has examined decades of study on what we know about the relationship between homework and scholastic achievement. He has proposed the “10-minute rule,” suggesting that daily homework be limited to 10 minutes per grade level. Thus, a 1st grader would do 10 minutes each day and a 4th grader, 40 minutes. The National Parent Teacher Association and the National Education Association both endorse this guideline, but it is not clear whether the recommended allotments include time for reading, which most teachers want children to do daily.

For middle-school students, Cooper and colleagues report that 90 minutes per day of homework is optimal for enhancing academic achievement, and for high schoolers, the ideal range is 90 minutes to two and a half hours per day. Beyond this threshold, more homework does not contribute to learning. For students enrolled in demanding Advanced Placement or honors courses, however, homework is likely to require significantly more time, leading to concerns over students’ health and well-being.

Notwithstanding media reports of parents revolting against the practice of homework, the vast majority of parents say they are highly satisfied with their children’s homework loads. The National Household Education Surveys Program recently found that between 70 and 83 percent of parents believed that the amount of homework their children had was “about right,” a result that held true regardless of social class, race/ethnicity, community size, level of education, and whether English was spoken at home.

Learning Beliefs Are Consequential

As noted above, developmentally appropriate homework can help children cultivate positive beliefs about learning. Decades of research have established that these beliefs predict the types of tasks students choose to pursue, their persistence in the face of challenge, and their academic achievement. Broadly, learning beliefs fall under the banner of achievement motivation, which is a constellation of cognitive, behavioral, and affective factors, including: the way a person perceives his or her abilities, goal-setting skills, expectation of success, the value the individual places on learning, and self-regulating behavior such as time-management skills. Positive or adaptive beliefs about learning serve as emotional and psychological protective factors for children, especially when they encounter difficulties or failure.

Motivation researcher Carol Dweck of Stanford University posits that children with a “growth mindset”—those who believe that ability is malleable—approach learning very differently than those with a “fixed mindset”—kids who believe ability cannot change. Those with a growth mindset view effort as the key to mastery. They see mistakes as helpful, persist even in the face of failure, prefer challenging over easy tasks, and do better in school than their peers who have a fixed mindset. In contrast, children with a fixed mindset view effort and mistakes as implicit condemnations of their abilities. Such children succumb easily to learned helplessness in the face of difficulty, and they gravitate toward tasks they know they can handle rather than more challenging ones.

Of course, learning beliefs do not develop in a vacuum. Studies have demonstrated that parents and teachers play a significant role in the development of positive beliefs and behaviors, and that homework is a key tool they can use to foster motivation and academic achievement.

Parents’ Beliefs and Actions Matter

It is well established that parental involvement in their children’s education promotes achievement motivation and success in school. Parents are their children’s first teachers, and their achievement-related beliefs have a profound influence on children’s developing perceptions of their own abilities, as well as their views on the value of learning and education.

Parents affect their children’s learning through the messages they send about education, whether by expressing interest in school activities and experiences, attending school events, helping with homework when they can, or exposing children to intellectually enriching experiences. Most parents view such engagement as part and parcel of their role. They also believe that doing homework fosters responsibility and organizational skills, and that doing well on homework tasks contributes to learning, even if children experience frustration from time to time.

Many parents provide support by establishing homework routines, eliminating distractions, communicating expectations, helping children manage their time, providing reassuring messages, and encouraging kids to be aware of the conditions under which they do their best work. These supports help foster the development of self-regulation, which is critical to school success.

Self-regulation involves a number of skills, such as the ability to monitor one’s performance and adjust strategies as a result of feedback; to evaluate one’s interests and realistically perceive one’s aptitude; and to work on a task autonomously. It also means learning how to structure one’s environment so that it’s conducive to learning, by, for example, minimizing distractions. As children move into higher grades, these skills and strategies help them organize, plan, and learn independently. This is precisely where parents make a demonstrable difference in students’ attitudes and approaches to homework.

Especially in the early grades, homework gives parents the opportunity to cultivate beliefs and behaviors that foster efficient study skills and academic resilience. Indeed, across age groups, there is a strong and positive relationship between homework completion and a variety of self-regulatory processes. However, the quality of parental help matters. Sometimes, well-intentioned parents can unwittingly undermine the development of children's positive learning beliefs and their achievement. Parents who maintain a positive outlook on homework and allow their children room to learn and struggle on their own, stepping in judiciously with informational feedback and hints, do their children a much better service than those who seek to control the learning process.

A recent study of 5th and 6th graders' perceptions of their parents' involvement with homework distinguished between supportive and intrusive help. The former included the belief that parents encouraged the children to try to find the right answer on their own before providing them with assistance, and when the child struggled, attempted to understand the source of the confusion. In contrast, the latter included the perception that parents provided unsolicited help, interfered when the children did their homework, and told them how to complete their assignments. Supportive help predicted higher achievement, while intrusive help was associated with lower achievement.

Parents' attitudes and emotions during homework time can support the development of positive attitudes and approaches in their children, which in turn are predictive of higher achievement. Children are more likely to focus on self-improvement during homework time and do better in school when their parents are oriented toward mastery. In contrast, if parents focus on how well children are doing relative to peers, kids tend to adopt learning goals that allow them to avoid challenge.

Across children's age groups, there is a strong and positive relationship between homework completion and self-regulatory processes.

Homework and Social Class

Social class is another important element in the homework dynamic. What is the homework experience like for families with limited time and resources? And what of affluent families, where resources are plenty but the pressures to succeed are great?

Etta Kralovec and John Buell, authors of *The End of Homework,* maintain that homework "punishes the poor," because lower-income parents may not be as well educated as their affluent counterparts and thus not as well equipped to help with homework. Poorer families also have fewer financial resources to devote to home computers, tutoring, and academic enrichment. The stresses of poverty—and work schedules—may impinge, and immigrant parents may face language barriers and an unfamiliarity with the school system and teachers' expectations.

Yet research shows that low-income parents who are unable to assist with homework are far from passive in their children's learning, and they do help foster scholastic performance. In fact, parental help with homework is *not* a necessary component for school success.

Brown University's Jin Li queried low-income Chinese American 9th graders' perceptions of their parents' engagement with their education. Students said their immigrant parents rarely engaged in activities that are known to foster academic achievement, such as monitoring homework, checking it for accuracy, or attending school meetings or events. Instead, parents of higher achievers built three social networks to support their children's learning. They designated "anchor" helpers both inside and outside the family who provided assistance; identified peer models for their children to emulate; and enlisted the assistance of extended kin to guide their children's educational socialization. In a related vein, a recent analysis of survey data showed that Asian and Latino 5th graders, relative to native-born peers, were more likely to turn to siblings than parents for homework help.

Further, research demonstrates that low-income parents, recognizing that they lack the time to be in the classroom or participate in school governance, view homework as a critical connection to their children's experiences in school. One study found that mothers enjoyed the routine and predictability of homework and used it as a way to demonstrate to children how to plan their time. Mothers organized homework as a family activity, with siblings doing homework together and older children reading to younger ones. In this way, homework was perceived as a collective practice wherein siblings could model effective habits and learn from one another.

In another recent study, researchers examined mathematics achievement in low-income 8th-grade Asian and Latino students. Help with homework was an advantage their mothers could not provide. They could, however, furnish structure (for example, by setting aside quiet time for homework completion), and it was this structure that most predicted high achievement. As the authors note, "It is . . . important to help [low-income] parents realize that they can still help their children get good grades in mathematics and succeed in school even if they do not know how to provide direct assistance with their child's mathematics homework."

The homework narrative at the other end of the socioeconomic continuum is altogether different. Media reports abound with examples of students, mostly in high school, carrying three or more hours of homework per night, a burden that can impair learning, motivation, and well-being. In affluent communities, students often experience intense pressure to cultivate a high-achieving profile that will be attractive to elite colleges. Heavy homework loads have been linked to unhealthy symptoms such as heightened stress, anxiety, physical complaints, and sleep disturbances. Like Allison's 6th grader mentioned earlier, many students can only tackle their homework after they do extracurricular activities, which are also seen as essential for the college résumé. Not surprisingly, many students in these communities are not deeply engaged in learning; rather, they speak of "doing school," as Stanford researcher Denise Pope has described, going through the motions necessary to excel, and undermining their physical and mental health in the process.

Fortunately, some national intervention initiatives, such as Challenge Success (co-founded by Pope), are heightening awareness of these problems. Interventions aimed at restoring balance in students' lives (in part, by reducing homework demands) have resulted in students reporting an increased sense of well-being, decreased stress and anxiety, and perceptions of greater support from teachers, with no decrease in achievement outcomes.

What is good for this small segment of students, however, is not necessarily good for the majority. As Jessica Lahey wrote in *Motherlode, a New York Times* parenting blog, "homework is a red herring" in the national conversation on education. "Some otherwise privileged children may have too much, but the real issue lies in places where there is too little. . . . We shouldn't forget that."

My colleagues and I analyzed interviews conducted with lower-income 9th graders (African American, Mexican American, and European American) from two Northern California high schools that at the time were among the lowest-achieving schools in the state. We found that these students consistently described receiving minimal homework—perhaps one or two worksheets or textbook pages, the occasional project, and 30 minutes of reading per night. Math was the only class in which they reported having homework each night. These students noted few consequences for not completing their homework.

Indeed, greatly reducing or eliminating homework would likely increase, not diminish, the achievement gap. As Harris M. Cooper has commented, those choosing to opt their children out of homework are operating from a place of advantage. Children in higher-income families benefit from many privileges, including exposure to a larger range of language at home that may align with the language of school, access to learning and cultural experiences, and many other forms of enrichment, such as tutoring and academic summer camps, all of which may be cost-prohibitive for lower-income families. But for the 21 percent of the school-age population who live in poverty—nearly 11 million students ages 5–17—homework is one tool that can help narrow the achievement gap.

Community and School Support

Often, community organizations and afterschool programs can step up to provide structure and services that students' need to succeed at homework. For example, Boys and Girls and 4-H clubs offer volunteer tutors as well as access to computer technology that students may not have at home. Many schools provide homework clubs or integrate homework into the afterschool program.

Home-school partnerships have succeeded in engaging parents with homework and significantly improving their children's academic achievement. For example, Joyce Epstein of Johns Hopkins University has developed the TIPS model (Teachers Involve Parents in Schoolwork), which embraces homework as an integral part of family time. TIPS is a teacher-designed interactive program in which children and a parent or family member each have a specific role in the homework scenario. For example, children might show the parent how to do a mathematics task on fractions, explaining their reasoning along the way and reviewing their thinking aloud if they are unsure.

Evaluations show that elementary and middle-school students in classrooms that have adopted TIPS complete more of their homework than do students in other classrooms. Both students and parent participants show more positive beliefs about learning mathematics, and TIPS students show significant gains in writing skills and report-card science grades, as well as higher mathematics scores on standardized tests.

Another study found that asking teachers to send text messages to parents about their children's missing homework resulted in increased parental monitoring of homework, consequences for missed assignments, and greater participation in parent-child conferences. Teachers reported fewer missed assignments and greater student effort in coursework, and math grades and GPA significantly improved.

Homework Quality Matters

Teachers favor homework for a number of reasons. They believe it fosters a sense of responsibility and promotes academic achievement. They note that homework provides valuable review and practice for students while giving teachers feedback on areas where students may need more support. Finally, teachers value homework as a way to keep parents connected to the school and their children's educational experiences.

While students, to say the least, may not always relish the idea of doing homework, by high school most come to believe there is a positive relationship between doing homework and doing well in school. Both higher and lower achievers lament "busywork" that doesn't promote learning. They crave high-quality, challenging assignments—and it is this kind of homework that has been associated with higher achievement.

What constitutes high-quality homework? Assignments that are developmentally appropriate and meaningful and that promote self-efficacy and self-regulation. Meaningful homework is authentic, allowing students to engage in solving problems with real-world relevance. More specifically, homework tasks should make efficient use of student time and have a clear purpose connected to what they are learning. An artistic rendition of a period in history that would take hours to complete can become instead a diary entry in the voice of an individual from that era. By allowing a measure of choice and autonomy in homework, teachers foster in their students a sense of ownership, which bolsters their investment in the work.

High-quality homework also fosters students' perceptions of their own competence by 1) focusing them on tasks they can accomplish without help; 2) differentiating tasks so as to allow struggling students to experience success; 3) providing suggested time frames rather than a fixed period of time in which a task should be completed; 4) delivering clearly and carefully explained directions; and 5) carefully modeling methods for attacking lengthy or complex tasks. Students whose teachers have trained them to adopt strategies such as goal setting, self-monitoring, and planning develop a number of personal assets—improved time management, increased self-efficacy, greater effort and interest, a desire for mastery, and a decrease in helplessness.

Excellence with Equity

Currently, the United States has the second-highest disparity between time spent on homework by students of low socioeconomic status and time spent by their more-affluent peers out of the 34 OECD-member nations participating in the 2012 Program for International Student Assessment (PISA) (see Figure 2). Noting that PISA studies have consistently found that spending more time on math homework strongly correlates with higher academic achievement, the report's authors suggest that the homework disparity may reflect lower teacher expectations for low-income students. If so, this is truly unfortunate. In and of itself, low socioeconomic status is not an impediment to academic achievement when appropriate parental, school, and community supports are deployed. As research makes clear, low-income parents support their children's learning in varied ways, not all of which involve direct assistance with schoolwork. Teachers can orient students and parents toward beliefs that foster positive attitudes toward learning. Indeed, where homework is concerned, a commitment to excellence with equity is both worthwhile and attainable.

In affluent communities, parents, teachers, and school districts might consider reexamining the meaning of academic excellence and placing more emphasis on leading a balanced and well-rounded life. The homework debate in the United States has been dominated by concerns over the health and well-being of such advantaged students. As legitimate as these worries are, it's important to avoid generalizing these children's experiences to those with fewer family resources. Reducing or eliminating homework, though it may be desirable in some advantaged communities, would deprive poorer children of a crucial and empowering learning experience. It would also eradicate a fertile opportunity to help close the achievement gap.

Janine Bempechat is clinical professor of human development at the Boston University Wheelock College of Education and Human Development.

An unabridged version of this article is available here.

For more, please see "The Top 20 Education Next *Articles of 2021."*

This article appeared in the Winter 2019 issue of Education Next*. Suggested citation format:*

Bempechat, J. (2019). The Case for (Quality) Homework: Why it improves learning, and how parents can help. *Education Next, 19*(1), 36-43.

Last updated October 10, 2018

Janine Bempechat

AUTHOR BIO

PREVIOUS ARTICLE

NEXT ARTICLE

We Recommend You Read

In the News: What's the Right Amount of Homework? Many Students Get Too Little, Brief Argues

by Education Next

In the News: Down With Homework, Say U.S. School Districts

by Education Next

In the News: Does Homework Really Help Students Learn?

by Education Next

THE JOURNAL

Spring 2022

Vol. 22, No. 2

Education Next is a journal of opinion and research about education policy.

SUBSCRIBE TODAY

NOT CONVINCED?

Exhibit 1.10

(page intentionally left blank)

Received: 30 May 2018 | Revised: 21 December 2018 | Accepted: 14 January 2019
DOI: 10.1111/medu.13830

TEACHING STRATEGIES

The stress paradox: How stress can be good for learning

Joy R. Rudland[1] | **Clinton Golding[2]** | **Tim J. Wilkinson[3]**

[1]Education Unit, University of Otago, Wellington, New Zealand
[2]Higher Education Development Centre, University of Otago, Dunedin, New Zealand
[3]Education Unit, University of Otago, Christchurch, New Zealand

Correspondence
Joy R. Rudland, Education Unit, University of Otago Wellington, 23A Mein Street, Newtown, Wellington 6242, New Zealand.
Email: joy.rudland@otago.ac.nz

Abstract

Context: This article explores the myth that stress is always bad for learning. The term "stress" has been narrowed by habitual use to equate with the negative outcome of *dis*tress; this article takes an alternative view that ultimately rejects the myth that demonises stress. The avoidance of distress is important, but a broader view of stress as something that can have either positive or negative outcomes is considered.

Proposal: We propose that stress is important for learning and stress-related growth. We explore the little-mentioned concept of *eu*stress (good stress) as a counter to the more familiar concept of distress. We further consider that the negative associations of stress may contribute to its negative impact. The impact of stress on learning should be deliberately and carefully considered. We offer a hypothetical learning journey that considers the cause of potential stress, a stressor, and how a stressor is moderated to result in stress that may influence learning either by positively challenging the learner or by functioning as a hindrance to learning.

Conclusions: In thinking more positively about stress, health professional educators may better support the student's learning journey.

1 | INTRODUCTION

This paper explores the common assumption that stress is bad for learning. We first describe how psychology, education and studies in occupation have used the broad term "stress" and then how this term has been narrowed by some to equate with an outcome and further narrowed to describe a negative outcome, distress. We also consider other research that suggests stress may be positive, with particular emphasis on how the customary framing of stress as inevitably bad masks the beneficial aspects of challenging situations. Ultimately rejecting the myth that demonises stress, we take a broader view of "stress" as something that can have either positive or negative outcomes. We reject the idea that stress is always to be avoided and propose a hypothetical learning pathway that positions stress as a necessary part of learning: a "stressor" prompts learning; moderation of the impact of the stressor occurs with a realisation of the stress experienced by the learner and finishes with how the stress is "actualised" in respect to the learning that has taken place and the associated effect of the learning. We propose a number of strategies that health professional educators may consider in order to enhance this learning pathway.

2 | HISTORICAL AND CURRENT USES OF "STRESS"

The term "stress" is used in both popular culture and the academic literature, notably that from psychology and education. In modern popular literature, as is evidenced by numerous stress-related self-help books, stress is often considered to be a sickness[1] and the term "stress" is frequently equated with an adverse outcome of an experience.[2]

Originally, the terms "stress" and "distress" were seen as two different concepts. The term "stress" was initially used in the contexts of metallurgy, physics and mathematics, or as a verb meaning "to give particular emphasis". By contrast, the term "distress" was used more frequently to describe biological manifestations such as respiratory and cardiac distress or digestive disorders. The concept of individuals being in distress, as opposed to biological systems, became evident in the 1950s with reference to, for example, people in distress, the distress of schoolchildren, and the distressed student.[3]

In health professional education today, the words "stress" and "distress" have come to be casually equated. The presence of stress tends to be portrayed as a hindrance to learning. Numerous articles in health professional education have reported "stress" in relation to a variety of stimuli including deficiency of knowledge,[4,5] lack of competence,[4,6] patient interaction,[7] questioning,[7] examinations, assessments and assignments,[4,7,8] and relationships with staff and teachers and the learning environment.[5,6] However, most of these studies look at what causes stress and often have a distinct bias or assumption that stress is bad and should be avoided. Consequently, the reduction of stress and the adoption of mental health strategies have been widely considered.[9]

It is true that medical students and doctors can perceive high levels of distress in education in comparison with other students[10,11] and professionals.[11] This can include "distress" associated with the bullying or humiliating of a learner.[12] The authors do not underestimate or diminish those stressors that equate with unreasonable or poor behaviour, maltreatment or unacceptable discrimination in learning environments. These behaviours have no place in health professional education and should not be appraised for anything other than what they are: unacceptable and damaging to learning.[13]

However, focusing only on distress may be limiting as it curtails recognition of the positive benefits of stress in health professional education. The term "eustress", coined by Selye,[14] means "a beneficial or healthy response to a stress, associated with positive feelings"[15] and is described as "an optimal amount of stress".[16] Literature about work-based stress similarly refers to an "optimal amount"[17] of stress. Eustress as a positive outcome of stress has been positively associated with high performance in sports[18] and work.[19] There is a clear distinction between "distress" and "eustress" as two different outcomes of a stressor.[20]

Various scales to measure levels of stress have been developed, but these quantify stress without considering the difference between positive and negative effects. Research looking at both distress and eustress, as they relate to health professionals, is relatively recent[21] and has included the development of a distress–eustress scale (termed "hassles and uplift"),[22] which opens the possibility of identifying and quantitatively framing stress in a positive manner.

There is much evidence that stress does not equal distress. In health professional education, increased perceived stress has been associated with increased levels of personal achievement in nursing students: those students with stress were more likely to go on to register as nurses, and less likely to burn out or to leave the course.[23] Importantly, the reporting of stress does not predict overall distress.[24] Some common sources of stress have been found to have no association with distress, such as difficulty and amount of work. Further, distressed and non-distressed students can experience the same stressors.[24] Stress has been linked to enhanced motivation, support-seeking behaviour and working harder.[25] Stress has been found to improve mental function, boost memory[26] and speed up brain processing.[27] It has also been found that a stressor after learning "emotionally laden content" can enhance memory.[26]

However, deleterious effects of stress on clinical reasoning have been reported[28] and from an educational perspective stress may be associated with narrowing attention[27,29] and reduced performance ability.[27] Although high levels of stress have been associated with poor academic performance, studies looking at subjective performance and stress have conflicting results.[30]

The Yerkes–Dodson law is often used to describe stress and performance. It proposes both a linear and an inverted U-shaped relationship between arousal and rates of learning, and was developed through observations of mice subjected to various electric shocks as they attempted to return to a nesting box.[31] The linear relationship related to rates of learning has generally been ignored,[32] but the inverted U gained huge popularity. The inverted U proposes that learning increases with physiological stimulation (stress) to a point at which the stress becomes too great and performance decreases. However, this law is unlikely to apply to human learning because the shocking of rodents performing simple physical tasks is not analogous to complex human psychology and context-related stress, and there is little empirical evidence in human learning to support this law. With reference to a single component of human learning, such as attention, it is clear that high levels of stress can both enhance and impair cognitive performance, which cannot be explained in terms of the Yerkes–Dodson law.[32] Although the Yerkes-Dodson law may be a myth, if applied to human learning, we mention it because it is popular and to guard against over-generalisation or -utilisation.

In summary, the original meaning of "stress" may have been contorted and as a consequence the potential values of stressors and the experience of stress have been undermined and diminished. Placing emphasis on distress in the context of exploring stress offers an incomplete picture.[21] Stress may be useful for learning, but we first need to repackage the potential value of stress.

3 | STRESS AS POSITIVE FOR LEARNING

To challenge the myth that "stress is bad", we start by providing clearer and more precise definitions for the educational context.

A stressor = a force that is applied.[33] Considering the learning context, this force is rephrased as a *challenge or learning expectation* (eg learning how to perform an invasive procedure, being questioned by a teacher, preparing for an assessment, or the learning environment).
Stress = a realisation by the learner that a stressor(s) exists (eg "I feel stress[ed] because of the examination"). This is a manifestation of the convergence of not only the stimulus (the stressor), but also of the learner and the broader environment. It may include a psychological, physiological or behavioural response to the stressor.
Outcome = affective disposition + learning. The affective disposition is generated and learning is achieved.
Distress = a negative affect as a result of stress.
Eustress = a positive affect as a result of stress.
Learning is either evident or not.

Put simply, stress results from the interpretation of a situation as challenging or hindering, and hence stress is different from the stressor. In

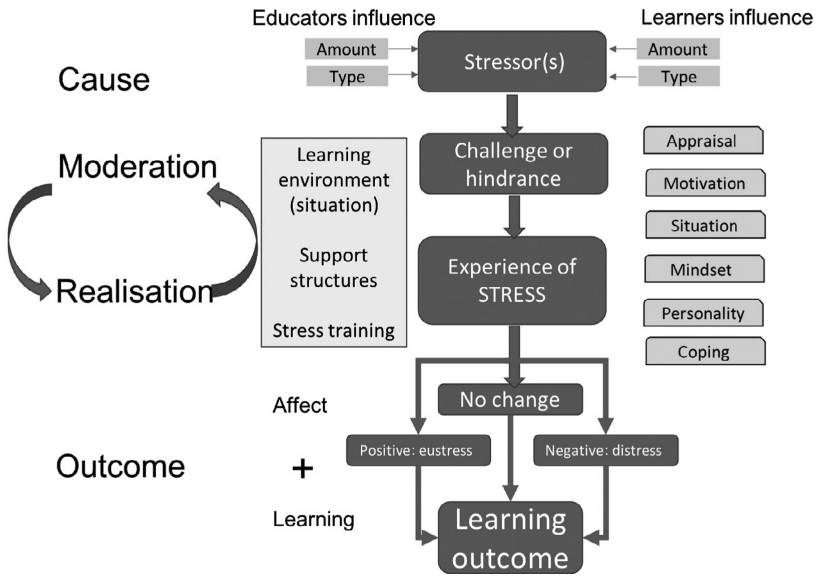

FIGURE 1 The pathway from the application of a learning stressor to its outcome

addition, stress may result in negative or positive affect, and may or may not result in learning.

This more precise definition makes it possible to explore the value of stress for learning. Our definition returns the term "stress" to its original intent and differentiates it from the outcome, the result of stress. In returning to this understanding of stress, we can revisit the influence stress may have on learning.

4 | A HYPOTHETICAL LEARNING PATHWAY INVOLVING STRESS

In this section, we present an account of learning that includes stressors and stress as parts of a learning journey. There are several models of stress, especially with respect to stress at work,[17] and we lean heavily on Spector's[34] compelling transactional model of stress, in which stress is a convergence between the environment and an individual.[16,34] We acknowledge the importance of the learning environment, specifically in the workplace, on learning (Figure 1).

4.1 | The stressor(s): the cause of stress in learning

It may be argued that learning has to start with a stressor(s). This may simply be the difference between what is known and what needs to be learned. It has been proposed that transformative change cannot occur without the stimulus of stress or crisis,[35] which results in what is termed "stress-related growth". As learning often occurs during an emotional episode, feeling positively stressed may be beneficial for learning.

It is useful to consider theories of learning at this juncture. Constructivism and transformative learning theories align well with the concept of a stressor as a necessity for the subsequent development of learning. Constructive learning theory requires a learner to be actively involved in the process of constructing meaning or knowledge,[36] whereas transformative learning results in a change in a person's viewpoint.[35] Both of these theories require the learner to engage with potential internal dissonance (a stressor).

A stressor may not necessarily be determined by the "amount" of learning to be done, but may refer to the "type" of learning expected. Challenge may come in many guises, such as in learning about a difficult topic, completing a skill or extracting important elements from a complex patient history. Both the student and the educator may be able to influence the amount and type of stressor applied, although this is frequently dictated by the educator and also the features of a work environment.

A stressor can be considered as an isolated challenge, but challenges may be combined and be additive. In thinking about the additive effects of stressors, cognitive load is apposite in conjuring the idea that an individual stressor may not elicit a negative response, but an accumulation of stressors may result in an intolerable level of stress.[37] The accumulation of stressors may be greater than the sum of the individual stressors and turn the effect of the stressors from being challenging to hindering.

There are many examples in health professions practice of contexts in which a particular type of stressor cannot and should not be avoided. For example, the learning of a clinical procedure (eg intramuscular injection) may be a stressor for students, and they may feel some stress when they first attempt to perform this procedure in a patient.

In health professional education, the clinical environment is also a stressor[5] because of the complex interactions between learning and patient care.[7] Trying to reduce stressors in the health professional learning environment may be futile, unrealistic and detrimental to learning growth.

4.2 | Moderation

The next step in the learning journey refers to how the learner moderates the stressor. Individuals will start to modify their response to a stressor immediately, whether consciously or unconsciously. How an individual interprets the influence of the stressor will influence the kind of stress he or she experiences, and whether learning takes place. In a biopsychosocial model, a stressor is interpreted as either challenging or threatening.[38] From an educational perspective, the value of stress can be usefully appraised according to whether it represents a hindrance or a challenge.[39-41] Challenge results from difficult demands that a person may feel confident about overcoming. In education, challenge is defined as being positive and necessary to the acquiring of new mental models.[41] Hindrance is defined as being negative and unsupportive for learning. These are the preferred terms as they focus on the reality of learning that can often be achieved even in the most difficult of circumstances.

Whether an individual interprets stress as challenging or hindering will be influenced by a number of factors, most notably:

Appraisal: response to a stressor is heavily influenced by how the stressor is *appraised* or evaluated, specifically how it is cognitively mediated.[16] It has been suggested that there is a primary appraisal related to the importance attached to the stressor and a secondary appraisal regarding whether an individual can "cope" with the stressor.[16]

Motivation of the learner: the motive of a learner to learn is crucial in influencing whether learning will occur.

Complexity of the situation: the response to a stressor may be influenced by the situation, whether the environment is busy or quiet, and whether the context involves many people or a one-to-one situation.

Mindset: a mindset or self-belief that being under stress is useful may have beneficial effects. A large study conducted in the USA found that the belief that stress was bad for the individual served as a self-fulfilling prophecy.[42] The mindset that decrees that stress is bad for the health was associated with poor health outcomes.[42] A more recent and education-focused study found that instructions that educated students about the adaptive benefits of stress resulted in improved performance by enhancing the students' perceptions of their ability to cope with the stressful testing situation.[43]

Personality traits: a personality type that is predisposed to negative or positive responses to stressors, such as one that is perfectionist,[44] subject to fear of failure or introverted, may influence a response to a stressor. Resilience, defined as the ability to "cope", is a personality trait that is particularly disposed to a positive response to stressors.

Coping strategies: coping strategies allow the learner to modify the feeling of stress. Individuals might have a "coping reservoir", which allows them to cope until the reservoir is depleted.[45] The consideration of coping strategies places a potentially unhelpful focus on the negative value of stress: it implies that stress is to be coped with rather than embraced. However, given that for many the nature of clinical work, irrespective of any learning occurring within it, can be depleting, recognition of the importance of strategies to fill the coping reservoir is to be applauded.

At this juncture, it would be reasonable to consider that a response to a stressor in the form of stress and the resultant outcome is totally within the ambit of the individual.[17] However, adopting this stance removes any responsibility from the part of the educator. Health professional educators have a role in supporting learners in interpreting stressors so that they result in eustress rather than distress.[17] Stressors may come from multiple sources, and performance and learning are facilitated if the stressors are related to the task or learning. The educator may help the learner to avoid stressors that are extraneous to the learning task because these will impair the achieving of the task.[30] Educators may also alter the complexity of the situation by determining whether a learning experience takes place in a pressured ward or in the more relaxed context of a teaching room, for example. Simulation may have an important role to play.

4.3 | Realisation

The moderation of the stressor by the learner may trigger a physical stress response. A physical stress response is an autonomic reaction to that stressor and is known as the first phase of "generalised adaptation syndrome".[46] The response may stimulate sympathetic nervous system activity and cortisol release; the heart rate may accelerate, and sweating may occur.

This response was initially described as the "fight or flight response". It is an acute response to danger. However, it is unclear whether a fight or flight interpretation is applicable to learning because learning is seldom dangerous. Firstly, it may not be that all autonomic reactions to stressors lead to a fight or flight response because the subjects involved in research in this area represented a biased sample: until 1995 only 17% of fight or flight research subjects were female.[47] In addition, to fight or flee may not be the only possible reactions: alternatives including "freeze" and "tend and befriend" responses have been proposed.[47] In addition, a stressor may not evoke any noticeable physiological response and even if physiological responses are evident, they do not necessarily need to be feared. Autonomic reactions to stress have been shown to improve performance at work in air traffic control. An increase in cortisol, a marker of being stressed, was found to correlate with higher peer ratings on competency and self-ratings on job satisfaction.[48]

Mindset can also influence how any physiological response is perceived. With a positive mindset towards stress, the perception of an increase in heart rate may be welcomed as beneficial instead of being viewed as detrimental.[2]

Educators have an important role in managing the learning environment to optimise the likelihood that a stressor will result in learning. Research in sport has found that manipulating the environment can buffer negative responses to stressors.[49] If an education institution offers support to learners, its learners will be more likely to experience stress positively. The value of support is well articulated through Dornan et al's work looking at experiential learning environments.[50] The potential that greater learning may be achieved when the learner is stressed or stretched can be aligned with the educational concept of the zone of proximal development, which refers to the difference between what is easy enough for a person to do on his or her own, and harder tasks that the same individual can complete only with support.[51] Support allows the learner to set harder and more demanding learning tasks or, in other words, to deal with greater stressors.[50,51]

4.4 | Actualisation: the outcome

The endpoint, the actualisation of the learning journey, is how the stressor, and the moderation and realisation of stress facilitate learning. Learning needs to have taken place at the conclusion of a learning journey; otherwise stress has no positive role in learning. Two distinct outcomes can be considered: one refers to how the learner feels about the learning experience (eustress or distress or nothing), and the other concerns whether learning has been achieved.

With respect to feelings of eustress or distress, it is not known whether distress always leads to limited learning and eustress to maximised learning. It is even unclear whether distress and eustress are on the same continuum or whether distress and eustress can be felt by the same individual simultaneously.[21] Whether a distressed individual can learn represents a quandary. Being distressed while learning does not seem to be either desirable or tenable. The state of eustress is desirable in savouring work[40] and likely in savouring the "work" of learning. Eustress has been linked to experiential learning.[21]

Although further research on the impact of eustress on learning is required, educators may have a role in ensuring stressors and learning are aligned and that stressors promote eustress rather than distress.

5 | CONCLUSIONS

In this article, we started with the myth that stress is bad for learning. This led to a need to better articulate what stress means with respect to learning. We propose that the term "stressor" be used as a noun to clarify a learning expectation that may be experienced by an individual. A learning stressor has the potential to be good or bad for learning. Along with the stressor, an individual's interpretation of and response to the stressor can make it either a positive challenge or a hindrance to learning.

The rejection of the myth allows us to propose that stressors and stress are important for learning and that we should be careful and deliberate in how we use stressors. Learning in a high-pressure workspace such as in clinical education has the potential to be stressful. However, how a stressor and the resultant stress are viewed and harnessed may improve how learners react to stress and, in consequence, influence the outcomes of learning and support the avoidance of distress as much as possible.

If we promote the notion that some stress may actually be beneficial for learning and consider the state of eustress as an important contributor to stress-related learning and growth, we can regard stress as beneficial. This article challenges the reader to avoid demonising stress and to be open to the possibility that stress can be beneficial for learning. Paradoxically, thinking about stress as being negative may contribute to its negative impact. Alternatively, we can focus on how we embrace and maximise the value of stress, even if we sometimes need to reduce stress.

Health professional educators have continuing roles in managing the type and amount of stressors experienced by learners and in more proactively helping learners see stressors as potentially stimulating eustress. This may be achieved by helping learners to think about the nature of stressors and to reframe negative mindsets. In addition, educators may enhance stress-related growth in the form of learning if they ensure that the learning environment is free from extraneous stressors and promote those stressors that are more likely to stimulate eustress.

In exploring the myth that "stress is always bad for learning", we hope to alert our readers to the possibility that stress is necessary for learning and, by doing so, begin a constructive dialogue about how we can maximise learning in conditions of stress.

ACKNOWLEDGEMENTS

None.

CONFLICT OF INTEREST

None.

AUTHOR CONTRIBUTION

All authors made contributions to the conception of the work. JRR drafted the first and subsequent drafts. CG and TJW contributed to the critical revision of the paper. All authors approved the final manuscript for submission and have agreed to be accountable for all aspects of the work.

ETHICAL APPROVAL

Not applicable.

ORCID

Joy R. Rudland https://orcid.org/0000-0002-7516-1728

REFERENCES

1. Evans W, Kelly B. Pre-registration diploma student nurse stress and coping measures. *Nurse Educ Today*. 2004;24 (6):473-482.
2. Crum AJ, Salovey P, Achor S. Rethinking stress: the role of mindsets in determining the stress response. *J Pers Soc Psychol*. 2013;104 (4):716-733.
3. Malleson N. The distressed student. *Lancet*. 1954;266 (6816):824-825.
4. Sheu S, Lin HS, Hwang SL. Perceived stress and physio-psycho-social status of nursing students during their initial period of clinical practice: the effect of coping behaviors. *Int J Nurs Stud*. 2002;39 (2):165-175.

5. Moczko TR, Bugaj TJ, Herzog W, Nikendei C. Perceived stress at transition to workplace: a qualitative interview study exploring final-year medical students' needs. *Adv Med Educ Pract*. 2016;7:15-27.
6. Gorostidi XZ, Huitzi Egilegor X, Jose Alberdi Erice M, Jose Uranga Iturriotz M, Eizmendi Garate I, Barandiaran Lasa M, Sanz Cascante X. Stress sources in nursing practice. Evolution during nursing training. *Nurse Educ Today*. 2007;27 (7):777-787.
7. Jimenez C, Navia-Osorio PM, Diaz CV. Stress and health in novice and experienced nursing students. *J Adv Nurs*. 2010;66 (2):442-455.
8. Burnard P, Edwards D, Bennett K, Thaibah H, Tothova V, Baldacchino D, Bara P, Mytevelli J. A comparative, longitudinal study of stress in student nurses in five countries: Albania, Brunei, the Czech Republic, Malta and Wales. *Nurse Educ Today*. 2008;28 (2):134-145.
9. Shiralkar MT, Harris TB, Eddins-Folensbee FF, Coverdale JH. A systematic review of stress-management programs for medical students. *Acad Psychiatry*. 2013;37 (3):158-164.
10. Dyrbye LN, Thomas MR, Shanafelt TD. Systematic review of depression, anxiety, and other indicators of psychological distress among US and Canadian medical students. *Acad Med*. 2006;81 (4):354-373.
11. beyondblue. National Mental Health Survey of Doctors and Medical Students. 2019. https://www.beyondblue.org.au/docs/default-source/research-project-files/bl1132-report---nmhdmss-full-report_web [Accessed 18 Feb 2019.]
12. Dahlin M, Joneborg N, Runeson B. Stress and depression among medical students: a cross-sectional study. *Med Educ*. 2005;39 (6):594-604.
13. Ye SJ, Yang R, Xiong Q, Yang Y, Zhou L, Gong Y, Li C, Ding Z, Ye G, Xiong Z. Acute stress enhances learning and memory by activating acid-sensing ion channels in rats. *Biochem Biophys Res Comm*. 2018;498 (4):1078-1084.
14. Selye H. *Stress without Distress*. Philadelphia, PA: JB Lippincott; 1974.
15. Nelson D, Simmons B. *Eustress: an Elusive Construct, an Engaging Pursuit*. Oxford, UK: Elsevier, JAI Press; 2004.
16. Lazarus RS, Folkman S. *Stress, Appraisal, and Coping*. London, UK: Springer Publishing; 1984.
17. Le Fevre M, Matheny J, Kolt G. Eustress, distress, and interpretation in occupational stress. *J Manag Psychol*. 2003;18 (7):726-744.
18. Bunting CJ, Little MJ, Tolson H, Jessup G. Physical fitness and eustress in the adventure activities of rock climbing and rappelling. *J Sports Med Phys Fitness*. 1986;26 (1):11-20.
19. Mukherjee D, Singh P. Converting of occupational stress to eustress: role of official hierarchy. *Int J Psychol*. 2008;43(3-4):797.
20. Suedfeld P. Reactions to societal trauma: distress and/or eustress. *Political Psychol*. 1997;18 (4):849-861.
21. Gibbons C, Dempster M, Moutray M. Stress and eustress in nursing students. *J Adv Nurs*. 2008;61 (3):282-290.
22. Gibbons C, Dempster M, Moutray M. Index of sources of stress in nursing students: a confirmatory factor analysis. *J Adv Nurs*. 2009;65 (5):1095-1102.
23. Deary IJ, Watson R, Hogston R. A longitudinal cohort study of burnout and attrition in nursing students. *J Adv Nurs*. 2003;43 (1):71-81.
24. Jones MC, Johnston DW. Distress, stress and coping in first-year student nurses. *J Adv Nurs*. 1997;26 (3):475-482.
25. Kaiseler M, Polman R, Nicholls A. Mental toughness, stress, stress appraisal, coping and coping effectiveness in sport. *Pers Individ Dif*. 2009;47 (7):728-733.
26. Cahill L, Gorski L, Le K. Enhanced human memory consolidation with post-learning stress: interaction with the degree of arousal at encoding. *Learn Mem*. 2003;10 (4):270-274.
27. Hancock PA, Warm JS. A dynamic model of stress and sustained attention. *Hum Factors*. 1989;31 (5):519-537.
28. Pottier P, Dejoie T, Hardouin JB, Le Loupp AG, Planchon B, Bonnaud A, LeBlanc VR. Effect of stress on clinical reasoning during simulated ambulatory consultations. *Med Teach*. 2013;35 (6):472-480.
29. Chajut E, Algom D. Selective attention improves under stress: implications for theories of social cognition. *J Pers Soc Psychol*. 2003;85 (2):231-248.
30. LeBlanc VR. The effects of acute stress on performance: implications for health professions education. *Acad Med*. 2009;84 (10 Suppl):S25-S33.
31. Yerkes RM, Dodson JD. The relation of strength of stimulus to rapidity of habit formation. *J Comp Neurol Psychol*. 1908;18 (5):459-482.
32. Christianson SA. Emotional stress and eyewitness memory: a critical review. *Psychol Bull*. 1992;112 (2):284-309.
33. Selye H. What is stress? *Metab Clin Exp*. 1956;5 (5):525-530.
34. Spector PE. A control theory of the job stress process. In: Cooper CL, ed. *Theories of Organizational Stress*. Oxford, UK: Oxford University Press; 1998;153-169.
35. Mahoney MJ, Marquis A. Integral constructivism and dynamic systems in psychotherapy processes. *Psychoanal Inq*. 2002;22 (5):794-813.
36. Cooper P. Paradigm shifts in designed instruction: from behaviorism to cognitivism to constructivism. *Educ Technol*. 1993;33 (5):12-19.
37. Rudland JR, Wilkinson TJ. When I say … stress. *Med Educ*. 2018;52 (7):692-693.
38. Blascovich J, Mendes W, Hunter S, Salomon K. Social 'facilitation' as challenge and threat. *J Pers Soc Psychol*. 1999;77 (1):68-77.
39. Cavanaugh M, Boswell W, Roehling M, Boudreau J. An empirical examination of self-reported work stress among US managers. *J Appl Psychol*. 2000;85 (1):65-74.
40. Hargrove MB, Nelson DL, Cooper CL. Generating eustress by challenging employees: helping people savor their work. *Organ Dyn*. 2013;42 (1):61-69.
41. Hattie J. *Visible Learning: A Synthesis of over 800 Meta-Analyses Relating to Achievement*. London, UK: Routledge; 2008.
42. Keller A, Litzelman K, Wisk LE, Maddox T, Cheng ER, Creswell PD, Witt WP. Does the perception that stress affects health matter? The association with health and mortality. *Health Psychol*. 2012;31 (5):677-684.
43. Jamieson JP, Mendes WB, Nock MK. Improving acute stress responses: the power of reappraisal. *Curr Dir Psychol Sci*. 2013;22 (1):51-56.
44. Kung CSJ, Chan CKY. Differential roles of positive and negative perfectionism in predicting occupational eustress and distress. *Pers Individ Dif*. 2014;58:76-81.
45. Dunn LB, Iglewicz A, Moutier C. A conceptual model of medical student well-being: promoting resilience and preventing burnout. *Acad Psychiatry*. 2008;32 (1):44-53.
46. Selye H. The alarm reaction, the general adaptation syndrome, and the role of stress and of the adaptive hormones in dental medicine. *Oral Surg Oral Med Oral Pathol*. 1954;7 (4):355-367.
47. Taylor SE, Klein LC, Lewis BP, Gruenewald TL, Gurung RA, Updegraff JA. Biobehavioral responses to stress in females: tend-and-befriend, not fight-or-flight. *Psychol Rev*. 2000;107 (3):411-429.
48. Rose RM, Jenkins CD, Hurst M, Herd JA, Hall RP. Endocrine activity in air traffic controllers at work. II. Biological, psychological and work correlates. *Psychoneuroendocrinology*. 1982;7 (2-3):113-123.
49. Hogue CM, Fry MD, Fry AC. The differential impact of motivational climate on adolescents' psychological and physiological stress responses. *Psychol Sport Exerc*. 2017;30:118-127.
50. Dornan T, Scherpbier A, Boshuizen H. Supporting medical students' workplace learning: experience-based learning (ExBL). *Clin Teach*. 2009;6 (3):167-171.
51. Wass R, Golding C. Sharpening a tool for teaching: the zone of proximal development. *Teach High Educ*. 2014;19 (6):671-684.

How to cite this article: Rudland JR, Golding C, Wilkinson TJ. The stress paradox: How stress can be good for learning. *Med Educ*. 2020;54:40–45. https://doi.org/10.1111/medu.13830

Exhibit 1.11

(page intentionally left blank)

Psychonomic Bulletin & Review (2020) 27:465–482
https://doi.org/10.3758/s13423-019-01694-7

THEORETICAL REVIEW

What explains the relationship between spatial and mathematical skills? A review of evidence from brain and behavior

Zachary Hawes[1] · Daniel Ansari[1]

Published online: 21 January 2020

Abstract
There is an emerging consensus that spatial thinking plays a fundamental role in how people conceive, express, and perform mathematics. However, the underlying nature of this relationship remains elusive. Questions remain as to how, why, and under what conditions spatial skills and mathematics are linked. This review paper addresses this gap. Through a review and synthesis of research in psychology, neuroscience, and education, we examine plausible mechanistic accounts for the oft-reported close, and potentially causal, relations between spatial and mathematical thought. More specifically, this review targets candidate mechanisms that link spatial visualization skills and basic numerical competencies. The four explanatory accounts we describe and critique include the: (1) *Spatial representation of numbers account*, (2) *shared neural processing account*, (3) *spatial modelling account*, and (4) *working memory account*. We propose that these mechanisms do not operate in isolation from one another, but in concert with one another to give rise to spatial-numerical associations. Moving from the theoretical to the practical, we end our review by considering the extent to which spatial visualization abilities are malleable and transferrable to numerical reasoning. Ultimately, this paper aims to provide a more coherent and mechanistic account of spatial-numerical relations in the hope that this information may (1) afford new insights into the uniquely human ability to learn, perform, and invent abstract mathematics, and (2) on a more practical level, prove useful in the assessment and design of effective mathematics curricula and intervention moving forward.

Keywords Spatial skills · Numerical skills · Spatial visualization · Mathematical cognition

Introduction

The mapping of numbers to space is central to how we operationalize, learn, and do mathematics. From a historical perspective, it is difficult, if not impossible, to sift through the major discoveries in mathematics without acknowledging the central importance placed on the mapping of numbers to space (Lakoff & Núñez, 2000). For example, the Pythagorean Theorem, the Cartesian coordinate system (mapping in general), triangular numbers, the real number line, and Cavalieri's principle are but a few famous examples of numerical-spatial mappings (Davis, 2015; Dehaene, 2011; Giaquinto, 2008; Hubbard, Piazza, Pinel, & Dehaene, 2009). More ubiquitous examples include the measurement of time and various other everyday quantities (e.g., cooking ingredients; Newcombe, Levine, & Mix, 2015). Mathematical instruments as well as measurement devices are in themselves a testament to the widespread application of mapping numbers to space. These examples include the abacus, number line, clock, and ruler. To flip through any mathematical textbook is to further reveal the intimate relations between numbers and space. Diagrams, graphs, and various other visual-spatial illustrations fill the pages and serve to communicate and improve mathematical understanding.

From these examples, it is clear that numbers and space interact in important ways. But how is it that these spatial-numerical associations come to be in the first place? What are the cognitive processes that underlie our uniquely human ability to derive the Pythagorean Theorem or to invent concepts and tools to measure the world around us? In this paper, we ask what role spatial abilities might play in mathematical reasoning. More specifically, we focus on the ways in which

✉ Zachary Hawes
zhawes@gmail.com

[1] Department of Psychology, Brain and Mind Institute, University of Western Ontario, Western Interdisciplinary Research Building, Pod A-10, 1151 Richmond Street North, London, ON N6A 5B7, Canada

spatial visualization might facilitate numerical reasoning skills, including the competencies related to basic number sense and operations. Our primary intent is to go beyond the question of whether spatial visualization and numerical abilities are linked and instead examine *why* they may be linked. The following quote not only speaks to this need, but also makes it clear why we should care about this area of study:

> *The relation between spatial ability and mathematics is so well established that it no longer makes sense to ask whether they are related. Rather, we need to know why the two are connected—the causal mechanisms and shared processes—for this relation to be fully leveraged by educators and clinicians* (Mix & Cheng, 2012, p. 206).

To address this gap in the literature, the following review presents four mechanistic accounts of why spatial visualization may be fundamentally linked to numerical reasoning. These four accounts include the: (1) *Spatial representation of numbers account*, (2) *shared neural processing account*, (3) *spatial modeling account*, and (4) *working memory account*. These accounts are based on a synthesis of literature spanning psychology, neuroscience, and education. They are not mutually exclusive. For example, there is considerable overlap between the spatial representation of numbers account and the shared neural processing account. The extent to which these various accounts are descriptions of the same underlying mechanism but vary according to discipline-specific epistemologies and research traditions, as well as different levels of analyses (behavioral vs. neural), is an important possibility to consider and one that we address in our *General discussion*. However, for ease of communication and in an attempt to best represent the research traditions in which these accounts originate from, we present them as separate mechanisms. In the end, it is our aim to provide insight and stimulate further questions as to when, why, and how spatial visualization and numerical abilities are linked.

We intentionally target spatial visualization skills in this review because this type of spatial thinking appears to be especially related to mathematical thinking (Mix & Cheng, 2012; Hawes, Moss, Caswell, Seo, & Ansari, 2019). For example, while there is little evidence (to date) to suggest that spatial navigation skills relate to mathematics abilities, there is well over a century of research linking spatial visualization and mathematics (Davis, 215; Galton, 1880; Mix & Cheng, 2012). Broadly defined here as the ability to generate, recall, maintain, and manipulate visual-spatial images and solutions (Lohman, 1996; see Fig. 1), spatial visualization has been reported to play a critical role in mathematical and scientific discovery and innovation. For example, the discovery of the structure of DNA, the Theory of Relativity, the Periodic Table, and the invention of the induction motor are all said to have been borne out of spatial visualization (Davis, 2015; Moss, Bruce, Caswell, Flynn, & Hawes, 2016; Newcombe, 2010). According to famed mathematician Jacques Hadamard (1945), mathematical discoveries first present themselves in the form of intuitions and visual-spatial imagery. Only then does one engage in the more arduous and time-consuming work of testing the veracity of one's imaginings through formal and symbolic logic. This theory is perhaps best articulated by Albert Einstein, who in a letter to Hadamard, wrote:

> *The words or language, as they are written or spoken, do not seem to play any role in my mechanism of thought. The physical entities which seem to serve as elements in thought are certain signs and more or less clear images which can be "voluntarily" reproduced and combined. …Conventional words or other signs have to be sought for laboriously only in a secondary stage, when the mentioned associative play is sufficiently established and can be reproduced at will* (Einstein, quoted in Hadamard, 1945, p. 142–143).

Critically, Einstein is not alone in describing his thought process in this way. Many other mathematicians and scientists, including Poincaré, van't Hoff, and Pasteur, have offered similar introspective accounts (Hadamard, 1945; Root-Bernstein, 1985). These anecdotal accounts provide important, but far from conclusive, accounts of the role(s) that spatial visualization might play in mathematical discovery. But what does the empirical evidence suggest? Further, and more to the point, what role do spatial visualization skills play in the learning and performance of school-based mathematics?

In terms of mathematical and scientific discovery and innovations, there is longitudinal support for strong predictive relations (Wai, Lubinski, & Benbow, 2009). For example, in a nationally representative sample of US high school students (N = 400,000), it was found that spatial visualization abilities predicted which students enjoyed, entered, and succeeded in STEM disciplines (science, technology, engineering, and mathematics), even after taking verbal and quantitative competencies into account (Wai, Lubinksi, & Benbow, 2009). Follow-up studies of this same sample further demonstrated that spatial visualization skills predicted creativity and innovation in the workplace, suggesting that there may be some truth to the anecdotal reports noted above (Kell, Lubinski, Benbow, & Steiger, 2013).

Consistent and robust correlations have been reported between spatial visualization skills and a breadth of mathematical tasks (Mix & Cheng, 2012). For example, spatial visualization skills have been linked to performance in geometry (Delgado & Prieto, 2004), algebra (Tolar, Lederberg, & Fletcher, 2009), numerical estimation (Tam, Wong, & Chan, 2019), word problems (Hegarty & Kozhevnikov, 1999),

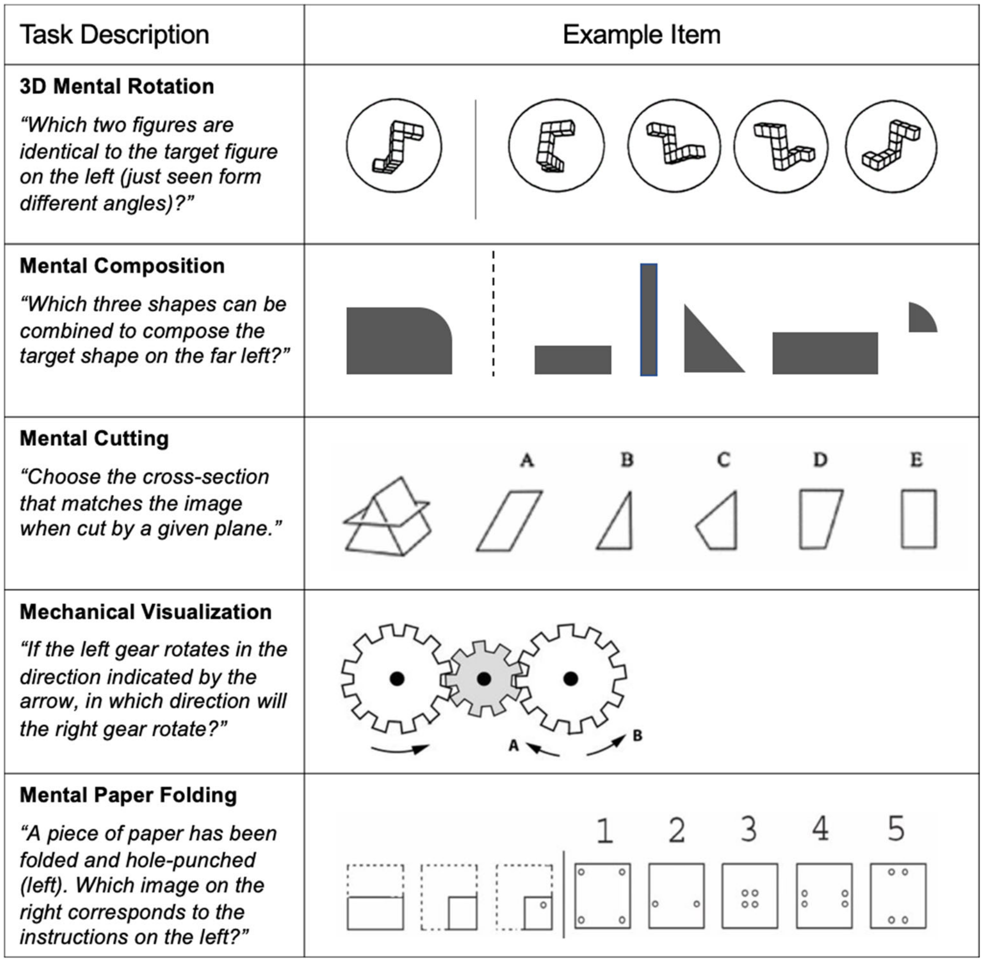

Fig. 1 Examples of measures used to capture individual differences in spatial visualization skills

mental arithmetic (Kyttälä & Lehto, 2008), and advanced mathematics (e.g., function theory, mathematical logic, computational mathematics; Wei, Yuan, Chen, & Zhou, 2012). Figure 1 presents a few examples of the types of measures that are typically used to capture individual differences in spatial visualization skills. In subsequent sections, we return to the question of what it is about this type of reasoning that might explain the consistent correlations with mathematics, generally, and with numerical reasoning more specifically.

As mentioned above, this review targets candidate mechanisms that link spatial visualization skills and basic numerical competencies. By basic numerical competencies we are referring to skills that relate to a basic understanding of number symbols and their various relations and applications (see Fig. 2 for examples). For example, tasks that require participants to compare and order numbers, perform arithmetic, and answer numerical word problems make up the sort basic numerical competencies we are referring to. Hereafter, the term mathematical and numerical reasoning will be used to refer this collection of tasks. The decision to specifically target numerical reasoning skills and not mathematics more generally was done for two reasons: First, in an effort constrain the literature search, and second, because the relationship between spatial visualization and numerical skills is not overtly obvious. While many branches of mathematics are inherently spatial, including geometry and measurement, the same cannot so easily be said of the most basic of mathematical entities and operations: numbers and arithmetic. Indeed, the question of why spatial visualization skills are linked to basic numerical competencies remains poorly understood. This review aims to provide insight into this question. We begin our review of the four accounts of why spatial visualization and numerical reasoning might be linked by considering the possibility that numbers may be characteristically spatial.

Spatial representation of numbers account

Numbers are the building blocks of mathematics: Their use omnipresent and fundamentally linked to almost all branches of mathematics. For this reason, any association between spatial processing and numbers is of potential critical importance in the effort to better understand the robust link between spatial skills and mathematics performance. As reviewed next,

Task Description	Example Item
Number Comparison *"As quickly but accurately as possible indicate the larger of two numbers"*	1 2 7 9
Number Ordering *"As quickly but as accurately as possible indicate whether or not the sequence of numbers is in correct (ascending) order or not."*	7 9 8 (✓ ×) 3 4 5 (✓ ×)
Number Line Estimation *"Indicate where the target number belongs on the number line."*	67; 0 ——— 100
Word Problems	*A balloon first rose 200 meters from the ground, then moved 100 meters to the east, then dropped 100 meters. It then traveled 50 meters to the east, and finally dropped straight to the ground. How far was the balloon from its original starting place?*
Arithmetical Operations	6 + 7, 8 + _ = 12, 8 – 3 + 5, etc.

Fig. 2 Examples of measures used to capture individual differences in numerical reasoning

there is a substantial body of research indicating that numbers may be represented spatially. According to a recent study on the subject, "*spatial processing of numbers has emerged as one of the basic properties of humans' mathematical thinking*" (Patro, Fischer, Nuerk, & Cress, 2016, pp. 126).

However, it remains unclear whether and to what extent spatial representations of number may confer any advantages to learning and doing mathematics. Moreover, and most germane to the purposes of the current review, it is not well understood what role higher-level spatial skills, namely spatial visualization skills, may play in the spatial representation of numbers.

The idea that numbers might be represented spatially has origins in Sir Francis Galton's mental imagery studies of the late 1800s (Galton, 1881). Galton provided the first evidence to suggest that numbers may be conceived as objects corresponding to specific positions in space:

> *Those who are able to visualize a numeral with a distinctness comparable to reality, and to behold it as if it were before their eyes, and not in some sort of dreamland, will define the direction in which it seems to lie, and the distance at which it appears to be. If they were looking at a ship on the horizon at the moment that the figure 6 happened to present itself to their minds, they could say whether the image lay to the left or right of the ship, and whether it was above or below the line of the horizon; they could always point to a definite spot in space, and say with more or less precision that that was the direction in which the image of the figure they were thinking of first appeared.* (1881, pp. 86)

Galton referred to such visualizations as number forms, noting that people's descriptions of such visualizations varied according to their visual-spatial properties, including differences in orientation, colour, and brightness. (e.g., see Fig. 3). Despite such differences, number forms were said to represent a reliable and stable trait within individuals.

Galton's studies on number forms is important because it provided the first evidence that people may represent numbers in a spatial format; most typically from left to right, akin to an actual number line. During the last several decades, considerable research efforts have followed up on this possibility through a wide assortment of empirical investigation. Perhaps the most influential study in this regard is Dehaene et al.'s (1993) original findings on the SNARC effect (Spatial-Numerical Association of Response Codes). In brief, the SNARC effect refers to the finding that people tend to automatically associate smaller numbers (e.g., 1, 2, 3) to the left side of space and larger numbers (e.g., 7, 8, 9) to the right side of space. People are faster and make fewer errors when

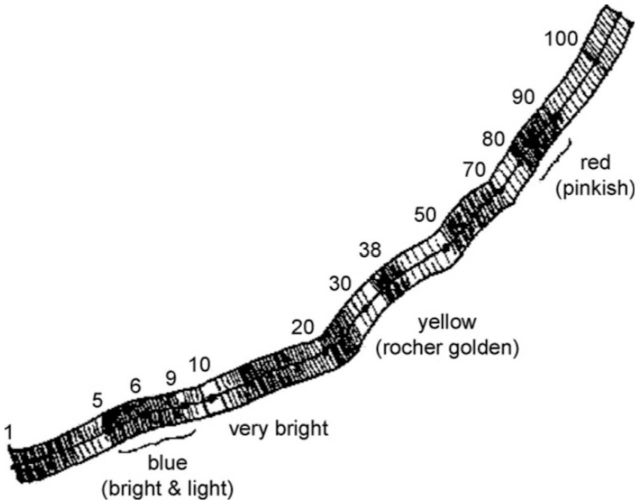

Fig. 3 An example of how one of the participants in Galton's study described their visualization of numbers

making parity judgments (i.e., determine whether a number is even or odd) when using the left hand to make judgements about small numbers and use the right hand to make judgements about larger numbers. This finding has been interpreted as evidence for the existence of a mental number line: A metaphor used to describe the tendency for individuals form Western cultures to conceive numbers as ordered magnitudes along a left-to-right axis. Indeed, the mental number line has been theorized to underlie a host of studies examining spatial-numerical associations (SNAs; e.g., see Toomarian & Hubbard, 2018). For example, line bisection tasks (Calabria & Rossetti, 2005), spatial attention tasks (Fischer & Fias, 2005), and even random number generation are but a few examples of tasks said to reveal spatial-numerical biases, interpreted as support for the presence of a "mental number line" (Loetscher, Bockisch, Nicholls, & Brugger, 2010). Arithmetic processing has also been suggested to induce automatic spatial-numerical biases (Knops, Viarouge, & Dehaene, 2009). For example, the operation-momentum effect refers to findings of left-right biases associated with addition and subtraction. Adult participants tend to overestimate answers to addition problems and underestimate answers to subtraction problems (McCrink, Dehaene, & Dehaene-Lambertz, 2007). Even when no calculation is required the mere presence of the operators themselves (i.e., + and −) have been found to influence left-right spatial biases (Mathieu et al., 2017). Importantly, evidence suggests that SNAs are mediated through cultural and educational practices. For example, the SNARC effect is reversed in cultures that read from right to left (Shaki, Fischer, & Petrusic, 2009). Taken together, there is considerable evidence to suggest that numerical thinking is related to spatial biases. These biases, in turn, have been taken as evidence of the "mental number line."

Critically, the mental number line has been posited to underlie both automatic/unconscious processing of numbers as well as more effortful/conscious processing of numbers (Fischer & Fias, 2005; Schneider et al., 2018; Toomarian & Hubbard, 2018). As we now demonstrate, this distinction has important implications in addressing the question of when and why spatial skills and numerical reasoning are related. While Galton's inquiries centered around conscious visualizations of number, the vast majority of studies on SNAs have examined automatic numerical-spatial biases. Research on the latter has revealed little evidence that SNAs are related to individual differences in numerical reasoning skills (Cipora, Patro, & Nuerk, 2015). Although a systematic review is needed to more fully investigate these relations, it is reasonable to conclude that automatic spatial-biases (as measured with the SNARC effect for example) have little influence on higher level numerical and mathematical processing. There is even some evidence to suggest that a negative association may exist. Practising mathematicians, for example, have been found to demonstrate weaker numerical-spatial biases compared to control subjects (Cipora et al., 2016). These findings stand in stark contrast to the research literature on intentional spatial-numerical associations (e.g., see Schneider et al., 2018).

For example, research on the number line estimation task reveals a consistent and reliable association between performance on the task and numerical reasoning. (Schneider et al., 2018). People who are more accurate at estimating where a given number belongs on a horizonal line flanked by two end points (e.g., 0 – 100; see Fig. 2), tend to also demonstrate better numerical and mathematical reasoning skills. Results of recent meta-analysis revealed an average correlation of .44 between number line task performance and mathematics (counting, arithmetic, school mathematics achievement) across the ages of 4–14 years (N = 10,576; Schneider et al., 2018). This effect size is considerably larger than the correlations that have been reported between other foundational numerical skills and mathematics achievement. For example, measures of symbolic number comparison – a widely accepted measure of numerical fluency – is estimated to share a .30 correlation with mathematics achievement (e.g., see Schneider et al., 2017). Moreover, to date, the most effective mathematics interventions have used the number line as the instructional tool used to enhance students' numerical reasoning (Fischer, Moeller, Bientzle, Cress, & Nuerk, 2011; Link, Moeller, Huber, Fischer, & Nuerk, 2013; Ramani & Siegler, 2008). Interestingly, number-line training studies are theorized to be effective because they lead to a more refined "mental number line" (Fischer et al., 2011; Siegler & Ramani, 2009).

Thus, in considering the above finings, we are left with an interesting paradox. Automatic/unconscious spatial-numerical associations do not appear to be related to individual differences in mathematics. On the contrary, intentional spatial-numerical associations appear to be strongly related to individual differences in mathematics. Moreover, both types of spatial-numerical associations – the unconscious and the conscious – are said to reflect the “mental number line.” What might explain this disconnect?

To gain insight into this question, we turn to the role that spatial visualization may play in first forming spatial-numerical associations. Several studies have now provided evidence that spatial visualization skills relate to improved number line performance, which in turn is related to improved arithmetic and mathematics performance (Gunderson, Ramirez, Beilock, & Levine, 2012; LeFevre, Jimenez Lira, Sowinski, Cankaya, Kamawar, & Skwarchuk, 2013; Tam, Wong, & Chan, 2019). In other words, linear numerical representations have been found to mediate relations between spatial and numerical reasoning. Other researchers have found that spatial visualization skills are positively correlated to automatic SNAs, including the SNARC effect (Viarouge, Hubbard, & McCandliss, 2014). It has been hypothesized that strong spatial visualization skills underlie a greater ease and fluency in which one can move up and down and carry out arithmetical operations along the mental number line (Viarouge, Hubbard, & McCandliss, 2014). However, this finding is somewhat at odds with the evidence viewed above. That is, if spatial visualization skills are linked to automatic SNAs, might we also expect automatic SNAs to relate to mathematics? Currently, it remains unclear whether, how, and why automatic SNAs mediate relations between spatial visualization and mathematics.

While it is easy to imagine the role that spatial visualization skills play in tasks that explicitly call upon the need to map numbers to space (e.g., the empty number line task), it is more difficult to imagine why spatial visualization skills are associated with automatic SNAs. One possibility is that automatic SNAs are an artefact of numerical-spatial relations formed earlier in development. That is, early in development, spatial visualization skills may help children to construct relations between space and number. Over time, children may internalize these spatial-numerical relations, a process that eventually gives rise to automatic numerical-spatial biases. An important question is whether spatial visualization skills are still related to automatic SNAs, once the “building process” is complete. While the study by Viarouge et al. (2014) suggests that the answer to this question is yes, this is the one and only study to directly address this question, to our knowledge. Moreover, even if follow-up research confirms relations between spatial visualization skills and automatic SNAs, we are still left with the question of why conscious SNAs but not automatic SNAs relate to mathematics.

One plausible explanation, related to the argument above, is that the intentional mapping of numbers to space involves a host of mathematical reasoning skills, including spatial and proportional reasoning (Barth & Paladino, 2011; Gunderson et al., 2012; Simms, Clayton, Cragg, Gilmore, & Johnson, 2016). Although more automatic mappings of number to space may also require these same skills, their influence on task performance may not be as paramount. Accordingly, the relation between spatial-numerical mappings and mathematics may be explained in part due to the extent that other mathematical relevant processes, including spatial visualization, are recruited during task execution. Said differently, mapping tasks that require higher levels of mathematical reasoning and precision are expected to share higher correlations with mathematical tasks that also require these same higher levels of precision and mathematical reasoning. This is a somewhat straightforward prediction, and, notably, one that aligns well with the spatial modeling account, but has yet to be investigated. As discussed in the next section, it is also possible that automatic SNAs are not as automatic as they appear, but rather constructed on the fly, within the confines of working memory and dependent on the specific task demands.

Critically, the mapping of numbers to space – by way of a mental number line – might be but *one* example in which spatial visualization skills are used to map and make sense of numerical-spatial relations (e.g., see Lakoff & Núñez, 2000; Marghetis, Núñez, & Bergen, 2014). As pointed out earlier, mathematics is full of examples in which numbers are mapped to space (e.g., geometric proofs, measurement, topology, etc.). Might spatial visualization skills play a more general role in mapping numbers, but also other mathematical entities and concepts, onto space? Indeed, as discussed earlier, the relationship between spatial visualization skills extends to a wide variety of mathematical tasks (Mix & Cheng, 2012). Moreover, numbers do not appear to be unique in their automatic association of left-right biases. For example, the SNARC effect has been extended and replicated with other ordered stimuli such as the days of week, months of the year, and letters of the alphabet (Gevers, Reynvoet, & Fias, 2003, 2004). Relatedly, the SNARC effect appears to be flexible and prone to priming effects. For example, Fischer et al. (2010) trained participants to view large numbers on the left and small numbers on the right and found evidence of a reversed SNARC effect (Fischer, Mills, & Shaki, 2010). Together, these findings suggest that the SNARC effect is (1) not limited to numbers, and (2) easily modulated by context. These findings have led to the hypothesis that the SNARC effect is indicative of context-dependent mappings between ordered stimuli (numbers, months, letters) and space.

Moreover, these findings challenge the long-held belief that numbers are inherently spatial and automatically associated with space. Instead, an alternative viewpoint has emerged, positing that numerical-spatial associations are constructed in working memory during task execution (van Dijck & Fias, 2011). Whether or not spatial visualization plays a role in this online constructive process remains to be studied. However, given the close link between spatial visualization skills and explicit numerical-spatial mappings (i.e., number line estimation tasks), spatial visualization skills may also facilitate more covert numerical-spatial mappings.

Taken together, questions remain regarding the extent to which numbers are automatically associated with space versus actively constructed on a moment-to-moment basis. Moreover, the role of spatial visualization in mapping numbers to space remains largely unknown. In the next section, we continue to expand on the central idea presented in this section – that is, spatial and numerical skills may be linked because numbers are represented spatially. While this section has revealed behavioral evidence in favor of a close coupling of numbers and space, the next section addresses questions about the neural mechanisms that underlie these relations.

Shared neural processing account

According to the shared neural networks account, spatial and numerical processing may be related because they rely on the same brain regions and make use of similar neural computations. The first indication that this may be the case came from neurological case studies. Individuals with damage to the parietal lobes were sometimes observed to demonstrate joint deficits in both spatial and numerical processing (Gerstmann, 1940; Holmes, 1918; Stengel, 1944). In fact, Gerstmann's Syndrome presents a rare but specific example of how damage to the parietal lobes (i.e., the left angular gyrus) is associated with impaired spatial and numerical reasoning. People with Gerstmann's Syndrome typically display a tetrad of symptoms including acalculia, left-right confusion, finger agnosia (difficulty identifying one's fingers), and dysgraphia (difficulty with writing) (Gerstmann, 1940). It has been suggested that these difficulties may be due to a more general deficit in the mental manipulation of visual-spatial images, including impaired mental rotation skills (e.g., see Mayer et al., 1999).

Research on patients with hemi-spatial neglect provides further evidence that space and number may depend on intact parietal lobes. Individuals with hemi-spatial neglect demonstrate an inability to attend to the contralesional portion of space (e.g., inability to attend to the left side of space when the lesion is in the right parietal lobe). This condition is associated with a skewed ability to indicate the mid-point of both real and imagined objects, but also the mid-point of numerical intervals (Bisiach & Luzatti, 1978; Zorzi et al., 2002). For example, Zorzi et al. (2002) asked right-brain-damaged patients to indicate the mid-point of two spoken numbers, such as "two" and "six." Presumably, due to an impaired mental number line, patients tended to overestimate the midpoint between two numbers as the interval between them increased (e.g., stating "five" as the midpoint between "two" and "six."). Taken together, neuropsychological case studies provide the earliest evidence that spatial and numerical processing may rely on common parietal cortex.

More recently, the advent of fMRI has given way to a host of follow-up investigations into the neural correlates of numerical and spatial thinking. This body of research points to the intraparietal sulcus (IPS) as the critical juncture in which numbers and space may interact (e.g., see Hawes, Sokolowski, Onoyne, & Ansari, 2019; Fig. 4). Indeed, it is now well established that the IPS and its neighboring regions play a critical role in reasoning about a variety of magnitudes, including non-symbolic quantities, space (size and shape), luminance, and even abstract notions such as number and time (see Kadosh, Lammertyn, & Izard, 2008; Hawes et al., 2019b; Walsh, 2003). Thus, there is evidence to suggest that *basic* spatial and numerical processes rely on common regions in and around the IPS.

There is also evidence that higher-level spatial skills, such as mental rotation, may also draw on these same parietal regions. For example, it has long been recognized that a central function of the parietal lobes is the performance of spatial transformations. Support for this can be seen in the results of a meta-analysis by Zacks (2008) on the neural correlates of mental rotation. Zacks found evidence found evidence to suggest that the IPS was the most robust and consistently activated brain region associated with mental rotation. Other spatial visualization processes, such as being able to compose/decompose and translate geometric shapes, have also been associated with activity in this region (Jordan, Heinze, Lutz, Kanowski, & Jäncke, 2001; Seydell-Greenwald, Ferrara, Chambers, Newport, & Landau, 2017). One reason that spatial and numerical reasoning may be linked is through shared processes related to mental transformations. According to Hubbard et al. (2009): "*parietal mechanisms that are thought to support spatial transformation might be ideally suited to support arithmetic transformations as well*" (2009, pp. 238). Indeed, this is an intriguing possibility and one that supports the neuronal re-cycling hypothesis.

According to the neuronal recycling hypothesis, numbers as well as other mathematical symbols and concepts may re-use the brain's neural resources that were originally specialized for interacting with the physical world (e.g., see Anderson, 2010, 2015; Dehaene & Cohen, 2007; Lakoff & Núñez, 2000; Marghetis, Núñez, & Bergen, 2014). In other words, numerical processing may co-opt or re-use the brain's more ancient and evolutionary adaptive spatial and sensorimotor systems, which originally served our abilities to interact with tools, objects, and locations in space (Dehaene et al., 2003; Johnson-Frey, 2004; Lakoff & Núñez, 2000). Marghetis et al. (2014) offer this summary of the neuronal re-cycling account: "*we may recycle the brain's spatial prowess to navigate the abstract mathematical world*" (pp. 1580). The neuronal recycling hypothesis has been used by many as an explanation for numerical-spatial biases observed through both behavioral and neuroimaging studies.

Taken together, there is compelling evidence that spatial and numerical processing are associated with overlapping

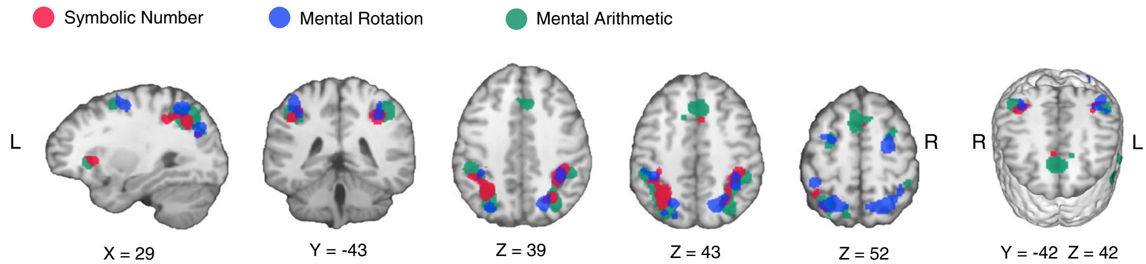

Fig. 4 Meta-analysis of fMRI studies examining brain regions associated with mental arithmetic (green), basic symbolic processes (red), and mental rotation (blue)

regions of the parietal cortex, namely in and around the IPS. However, there are also some notable gaps in the literature. One such gap is the emphasis placed on uncovering how *basic* spatial processes (e.g., comparing line lengths) relate to *basic* numerical processes (e.g., comparing Arabic digits; e.g., see Sokolowski, Fias, Mousa, & Ansari, 2017). To date, research on higher-level spatial skills (i.e., those that require spatial transformation, such as mental rotation) have been studied in isolation from neuroimaging studies of numerical cognition (but see Hawes et al., 2019b). So, although there is good evidence to suggest that higher-level spatial skills also rely on processes associated with the IPS, we do not yet have any direct evidence (i.e., from a single study) for this correlation. However, this is a critical gap in the literature for reasons discussed earlier. While there is robust evidence for relations between spatial visualization skills and numerical and mathematical performance, there is little evidence that spatial representations of number relate to individual differences in numerical and mathematical performance. Thus, when it comes to better understanding individual differences in mathematics performance, much can be gained by studying the neural relations between spatial skills proper and numerical and mathematical reasoning.

Spatial modeling account

According to the spatial modeling account, spatial visualization is related to numerical reasoning because it provides a "mental blackboard" of which numerical relations and operations can be modeled and visualized. More specifically, spatial visualization has been posited to play a critical role in how one organizes, models, and ultimately conceptualizes *novel* mathematical problems (Ackerman, 1988; Mix et al., 2016; Uttal & Cohen, 2012). Although there may be little to no need to model familiar mathematical content, such as memorized arithmetic facts, the visualization process may prove beneficial when confronted with novel mathematical content, such as arithmetic questions that require multi-step calculations. Moreover, the spatial modeling account functions to bridge past, present, and future knowledge states. For example, to solve the question 58 + 63, one might use prior knowledge of arithmetic facts to arrive at a previously unknown arithmetic fact (e.g., reason that 50 + 60 = 110 and 8 + 3 = 11; therefore, the solution is 110 + 11 = 121). To do this – bridge prior knowledge with newly created knowledge – one must also maintain the problem and interim solutions in mind. Whether or not these same functions might just as easily be ascribed to a working memory account is an important question and one we further address below.

Arguably, the most impressive feature of the spatial modeling account, but also perhaps its Achilles heel when it comes to empirical study, is that there are few, if any, limitations on the type of mathematical content that can be modeled by way of spatial visualization. Indeed, spatial visualization processes provide a space in which one can move back and forth between a multitude of representations; between the concrete and the abstract, the symbolic and the nonsymbolic, the real and the imagined, and static and dynamic representations (Antonietti, 1999). In short, there appear to be few limitations on the types of mathematical relations that can be modeled through visualizations. It is for this reason that it can be difficult to empirically investigate the spatial modeling account. How does one reveal the specific type of spatial modeling that occurs in the "mind's eye" of any given individual? Are some types of spatial modeling more conducive to effective mathematical reasoning than others?

One promising approach to these questions comes from studying how children model solutions to mathematical word problems. For example, Hegarty and Kozhevnikov (1999) presented children with the following word problem:

> *"A balloon first rose 200 meters from the ground, then moved 100 meters to the east, then dropped 100 meters. It then traveled 50 meters to the east, and finally dropped straight to the ground. How far was the balloon from its original starting place?"*

Children's accompanying drawings to the problem revealed key insights and differences into how children modeled/visualized the problem. While some children's drawings were literal representations of the problem, others were more abstract and contained only the relevant mathematical details needed to answer the question. Based on these differences, children's drawings were categorized as either pictorial (more literal in representation) or visual-schematic (more abstract in representation; emphasis on relevant numerical-spatial relations; see Fig. 5 for an example). Children who produced visual-schematic representations were more likely to arrive at the correct solution. Moreover, children who produced visual-schematic representations were also found to demonstrate significantly higher spatial visualization skills. Several studies have since replicated this finding (see Boonen, van der Schoot, van Wesel, de Vries, & Jolles, 2013; Boonen, van Wesel, Jolles, & van der Schoot, 2014). Taken together, these studies suggest that spatial visualization skills may indeed help learners to better model mathematical relations, which in turn, may lead to improved mathematical performance.

In the above studies on word problems, it appears best to create mental models of only the relevant mathematical details. However, the question of what to model is likely task/question specific. For some maths problems, it is not so much about "doing away" with irrelevant details, but about retaining, manipulating, and forming new relations with the information given. For example, take missing term problems, such as 5 + __ = 7. It has been suggested that one of the ways in which children come to develop fluency with such questions is through the ability to re-structure (re-model) the problem. So, instead of 5 + __ = 7, the learner might transform the question into the more familiar form, ___ = 7-5. What role might spatial visualization skills play in this process? To investigate this question, Cheng and Mix (2014) carried out a randomized controlled trial with 6- to 8-year-olds. Half the children were assigned to mental rotation training condition and the other half were assigned to an active control group. Compared to the control group, children in the mental rotation group demonstrated significant gains on the missing term problems. Consistent with the spatial modeling account, the authors suggested that gains on the missing term problems may have a resulted from an improved ability to re-model the problems into an easier format. This study provided the first causal evidence that spatial visualization training may transfer to mathematics. However, a recent follow-up study by Hawes et al. (2015) failed to replicate this finding. It is clear that more research is needed before causal claims can be made about the generalizability of spatial training to mathematics. In moving forward, such efforts should also try to more specifically address the mechanism of transfer. For example, what evidence is there that the changes in mathematics occur because of the effect that spatial training has on the way the problems are modelled? Insights into this question are needed in order to test the validity and make causal claims about the spatial modeling account.

As mentioned earlier, one of the predictions of the spatial modeling account is that spatial modeling is most used when dealing with novel versus familiar mathematical content. There is some evidence that this may be the case. To test this possibility, Mix et al. (2016) examined the relation between spatial skills, including spatial visualization, and novel and familiar mathematical content. Their results suggested that spatial skills were most closely related to novel mathematical problems. A follow-up study by Hawes et al. (2019a) provides additional insights into this issue. Using latent variable analyses, it was found that spatial visualization skills were highly correlated to both basic numerical skills as well as more advanced numerical skills (e.g., applied number problems, number operations). However, the relations between spatial visualization skills and higher-level numerical skills were much stronger than relations between spatial visualization skills and basic numerical skills. These studies provide some important preliminary support for the spatial modeling account. However, these studies do not provide any direct evidence that spatial visualization is differentially used as a function of problem familiarity or difficulty.

It is important to note that the spatial modeling account overlaps with other theories of numerical and mathematical cognition. In particular, it bears close resemblance with grounded and embodied accounts of mathematical cognition

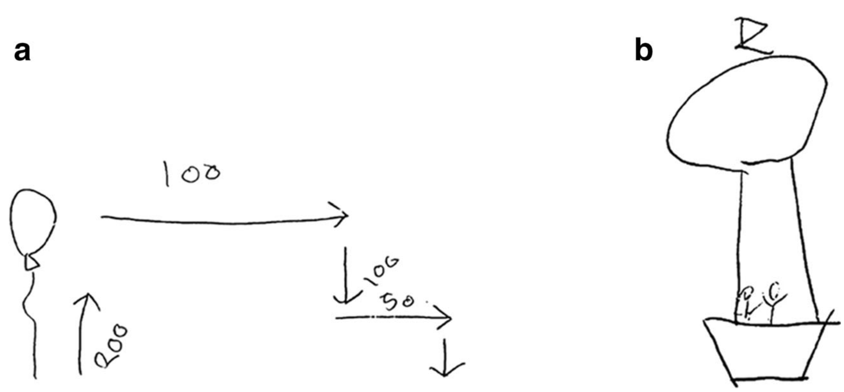

Fig. 5 An example of a visual-schematic representation (**A**) vs. a pictorial representation (**B**)

(Lourenco et al., 2018). According to these perspectives, mathematical thought is grounded in our everyday sensory and bodily experiences (Anderson, 2010, 2015; Lakoff & Núñez, 2000; Marghetis, Núñez, & Bergen, 2014). It is through engaging with metaphors, mental imagery, and simulated actions that mathematics becomes meaningful, and ironically, "groundless" (e.g., see Lakoff & Núñez, 2000). This view contrasts with the perspective that mathematics is largely independent of sensorimotor experiences and instead is a function of symbolic amodal thought (e.g., see Barsalou, 2008). Most relevant to the spatial modeling account is the role that mental simulation has been hypothesized to play in cognition in general, and in mathematics, in particular (Anderson, 2016; Barsalou, 2008; Huttenlocher, Jordan, & Levine, 1994). Indeed, mental simulation and mental modeling are alike in that they describe mental processes related to the re-enactment of sensorimotor experiences (e.g., mental imagery) in the service of a future goal (e.g., arriving at the correct solution to a word problem). The following provides an apt summary of the grounded cognition account, including clear parallels with mental simulation and the spatial modeling account:

> *Operations with some of the objects in mental models are like operations with physical objects. In reasoning about these objects, the person mentally moves about on them or in them, combines them, changes their sizes and shapes, and performs other operations like those that can be formed on objects in the physical world* (Greeno, 1991, pp. 178).

To be clear, the spatial modeling account is a more specific instantiation of mental simulation; one that is confined to the discipline of mathematics and deals explicitly with spatial relations. The above quote speaks to the "neuronal recycling" hypothesis mentioned earlier, offering additional insights into why space and number might both heavily recruit bilateral regions in and around the IPS. It is possible that numbers and various other mathematical concepts are processed in ways similar to the planning and actions associated with our handling of everyday objects. This point perhaps speaks to the common practice amongst mathematicians to refer to numbers as well as other mathematical concepts and abstractions as objects (Font, Godino, & Gallardo, 2013). A better understanding of why and how mathematicians come to view various concepts as objects may prove useful in better understanding the spatial modeling account as well as the development of mathematical expertise. For example, decades of research on human learning and memory indicate that objects are more easily remembered and expressed than abstract concepts (Paivio, 1983, 2013). Might this same finding apply to the realm of mathematics? Rather than dealing with isolated fragments of mathematical procedures carried out in a step-by-by fashion, presumably under the control of a more verbally mediated cognitive system, might the mathematical mind operate with greater ease and efficiency when dealing with holistic and object-like mental models? At the moment, research into these as well as other related questions concerning the spatial modeling account remain scarce. As such, the spatial modeling account remains a largely speculative account of why spatial visualization and numerical reasoning are so often linked.

Working memory account

Another way in which spatial visualization and numerical skills may be related is through another variable which shares relations with performance in both of these areas. For example, it is possible that spatial visualization skills are essentially a proxy for other cognitively demanding skills, such as executive function skills, working memory, and general intelligence. Visual-spatial working memory (VSWM), in particular, may explain the relations between spatial visualization and numerical skills. In this section, we review the evidence for and against this proposal.

Research to date suggests that both spatial visualization skills and VSWM are strongly related to numerical reasoning. For example, performance on spatial visualization tasks, such as mental rotation, have been linked to basic measures of numerical competencies, including arithmetic, number comparison, and number line estimation. Similarly, VSWM has also been found to explain similar amounts of variance in these same measures. Furthermore, there is evidence of close relations between all three of these variables – VSWM, spatial visualization, and numerical reasoning – when measured concurrently in the same studies (Alloway & Passolunghi, 2011; DeStefano & LeFevre, 2004; Hawes et al., 2019a; Kaufman, 2007; Kyllonen & Christal, 1990; Kyttälä et al., 2003; Li & Geary, 2017; Mix et al., 2016). Together, these findings question the extent to which spatial visualization and VSWM skills make unique contributions to numerical abilities.

It has been suggested that poor spatial abilities are a result of low VSWM. For example, several researchers have demonstrated notable differences in people of low- versus high-spatial abilities in their abilities to form, maintain, and transform visual-spatial representations (Carpenter & Just, 1986; Just & Carpenter, 1985; Lohman, 1988). Carpenter and Just (1986) concluded that "*a general characterization...is that low spatial subjects have difficulty maintaining a spatial representation while performing transformations*" (p. 236). That is, individuals with low-spatial abilities tend to "lose" information as they engage in the act of spatial transformation. For example, when mentally rotating cube figures, individuals with low-spatial abilities often lose "sight" of the mental image and require multiple attempts at rotation (Carpenter & Just, 1986; Lohman, 1988). Against this background, researchers have attributed individual differences in spatial visualization as primarily due to differences in working memory (e.g., see Hegarty & Waller, 2005).

Evidence to suggest that spatial visualization skills and VSWM are not as related as suggested above comes from three separate bodies of research: factor analyses, research on sex differences, and training studies. Studies from factor analytic studies suggest that VSWM, spatial visualization, and executive functions represent distinct cognitive constructs (i.e., latent variables; Hawes et al., 2019a; Miyake et al., 2001). Moreover, Hawes et al., 2019a demonstrated spatial visualization and numerical skills (both basic and advanced) not only represent distinct constructs, but that the relations between the two could not be explained by general intelligence or executive functions, including measures of VSWM. Lastly, Mix et al. (2016) found evidence that in sixth grade, VSWM shared stronger cross-loadings with a general mathematics factor compared to measures of spatial visualization, which were more associated with a general spatial factor. Together, these findings indicate that spatial visualization and VSWM represent separable cognitive factors and share differential relations with numerical and mathematics performance.

Further evidence that spatial visualization and VSWM are separable constructs can be gleaned from findings of reliable sex differences on measures of spatial visualization but not VSWM (Halpern et al., 2007).[1] Beginning by about the age of 10 years, males tend to outperform females on measures of mental rotation, with estimated effects sizes ranging from .9 – 1.0 (Halpern et al., 2007; Titze, Jansen, & Heil, 2010). Importantly, sex differences are not confined to mental rotation tasks but also emerge on other spatial visualization tasks, including mental paper folding tasks (Halpern et al., 2007). Findings of sex differences in spatial visualization skills, but not VSWM, further suggests that these two aspects of visual-spatial processing may represent distinct constructs.

Training studies provide further evidence that VSWM and spatial skills behave and operate in unique ways. Although the effects of VSWM training are hotly debated and there is little evidence that training generalizes to other untrained tasks (e.g., mathematics; Redick, Shipstead, Wiemers, Melby-Lervåg, & Hulme, 2015), a different picture has emerged with respect to spatial training. A recent meta-analysis of 217 spatial training studies by Uttal et al. (2013) indicates that spatial thinking can be improved in people of all ages and through a wide assortment of training approaches (e.g., course work, task-based training, video games). Furthermore, the researchers concluded that although further evidence is still required, it appears as though the effects of spatial training transfer to a variety of novel and untrained spatial tasks. In subsequent sections, we return to the topic of spatial training and the extent to which spatial training transfers to numerical reasoning. The take away point in this section, however, is that compared to VSWM, spatial visualization skill appears to represent a more flexible and adaptive cognitive system, providing further insight into the separability of VSWM and spatial skills.

At this point, it is worth returning to the question at hand: Does VSWM explain the relationship between spatial visualization skills and numerical/mathematical abilities? Based on the available evidence, there are reasons to suspect that (1) spatial visualization and VSWM are separable constructs, and (2) that each share independent pathways with numerical skills. An important follow-up question is why VSWM and spatial visualization skills may differentially contribute to numerical and mathematical learning and performance.

One proposal is that VSWM and spatial visualization differ according to the cognitive demands placed on the need to "recall" versus "generate" visual-spatial information. For example, at a measurement level, most VSWM measures primarily require participants to *recall*, maintain, and (sometimes) manipulate visual-spatial information. Most spatial visualization measures, on the other hand, require participants to perceive, maintain, manipulate, and *generate* visual-spatial solutions. Thus, the shared need to maintain and manipulate visual-spatial information may explain the previously reported correlations between VSWM and spatial visualization. However, the differences in task requirements might be one reason to predict differential relations with numerical performance. While VSWM skills may play a greater role in numerical tasks that emphasize the need to recall and maintain information (e.g., basic arithmetic), spatial visualization skills may play a greater role in numerical tasks that emphasize the need to generate novel solutions (e.g., word problems, applied problems). Notably, this prediction supports the spatial modeling account discussed earlier. Spatial visualization skills are predicted to be especially useful, even more so than VSWM, on problems that require the modeling and generation of problem solutions. Future research is needed to formally test this hypothesis.

Discussion

To this point, the relationship between spatial ability and mathematics has been well-studied but scarcely understood. The purpose of this review was to shed light on this issue by reviewing the literature in search of potential mechanisms that might explain the historically tight relations between spatial

[1] It should be recognized that this argument also applies to relations between spatial visualization and numerical reasoning. Although sex differences are frequently observed on measures of spatial visualization (namely mental rotation), sex differences do not regularly occur on measures of numerical reasoning (e.g., see Hutchison, Lyons, & Ansari, 2019; Kersey, Braham, Csumitta, Libertus, & Cantlon, 2018). This finding provides an additional constraint to consider in the attempt to disentangle the link between spatial visualization and numerical reasoning.

and numerical reasoning. More specifically, this review targeted the ways in which spatial visualization might be linked to numerical reasoning. Based on comprehensive review of research from psychology, neuroscience, and education, four potential mechanisms were identified: (1) *Spatial representation of numbers account*, (2) *shared neural processing account*, (3) *spatial modeling account*, and (4) *working memory account*. In the following section, a brief summary of each account is provided. We then engage in a more thorough discussion of the limitations as well as potential for improving spatial-numerical relations.

A summary of the four accounts

In brief, the *spatial representation of numbers account* suggests that numbers and their various relations are represented along a "mental number line." In turn, the precision of one's mental number line has been posited to play an important role in performing a host of numerical reasoning tasks, including comparing, ordering, and operating on numbers. A small body of research suggest that spatial visualization skills play a fundamental role in the learning and formation of numerical representations. The *shared neuronal processing account* suggests that numbers and space are linked through shared underlying neuroanatomical substrate. According to the neuronal recycling theory, numerical reasoning capacities re-use the same neuronal resources that were originally (evolutionarily speaking) deployed for spatial reasoning, including spatial visualization. The *spatial modeling account* places much more emphasis on spatial visualization as a more general mechanism used to model, organize, and simulate a wide variety of numerical concepts. This account is closely connected to other theories of mathematical reasoning, including grounded, embodied, and the aforementioned neuronal recycling accounts. A common feature of these theories is the use of the visual-spatial imagination to act upon mathematical objects, including numbers, in ways not unlike the ways we experience and use objects in the real world. Lastly, the *working memory account* calls into question unique relations between spatial visualization and numerical reasoning skills. Instead, the link may have its roots in individual differences in visual-spatial working memory (VSWM). However, evidence to date suggest that these two constructs are not one and the same and make independent contributions to numerical and mathematical performance. Moving forward, it will be important to continue to study the ways in which spatial visualization represents a unique construct as well as the ways in which it interacts with other cognitive systems.

An integrated description of the four accounts

The extent to which these various accounts are descriptions of the same underlying mechanism but in different forms and at different levels of analysis is an important question. For example, it is possible that one of the ways in which numbers become represented spatially is through the active processes of spatial modeling (e.g., visualizing a number line to reason about numerical relations). From a biological perspective, it could be that the IPS and closely associated regions provide the necessary neuronal networks to carry out these modeling and transformational processes. Moreover, even when the spatial modeling of numerical concepts no longer serves the individual (i.e., the concepts at hand have become automatized more or less), these same neural substrates may continue to underlie both numerical and spatial processes (e.g., see Hawes et al., 2019b). This may occur despite an independence in function. If we assume that spatial visualization is a relatively stable trait, then we should expect to see lasting correlations between spatial visualization and numerical skills even when spatial visualization no longer serves a purpose in one's semantic understanding/representation of number. In other words, spatial and numerical processes may continue to be correlated, both neurally and behaviorally, long after they have become conceptually divorced from one another. This relation may remain because of individual differences in spatial visualization skills that once helped give rise to conceptual mappings between numbers and space. This integrated account may explain why we continue to see correlations between spatial visualization skills and basic numerical competencies into adulthood. It might also explain why we see relations between intentional numerical-spatial mappings (e.g., as measured with the number line task) and mathematics (Schneider et al., 2018), but mixed evidence for relations between automatic numerical-spatial mappings (i.e., SNARC) and mathematics (Cipora, Patro, & Nuerk, 2015).

From this example it can be seen how biology and behavior interact in complex ways to give rise to potentially dynamic and ever-changing numerical-spatial relations. To what extent does genetically endowed neuroanatomical structures influence one's abilities to visualize numerical-spatial relations? To what extent is spatial visualization malleable and transferable to numerical reasoning? These are important questions; the answers of which may help to more fully understand the interplay that may exist between the various accounts of space-number relations.

Biological considerations of the four accounts

An interesting question concerns the extent to which one's spatial visualization abilities are constrained by genetic and corresponding neuroanatomy. Results from a meta-analysis of twin studies (N=18,296 monozygotic twins; N=23,327 dizygotic twins) suggest that spatial visualization abilities are largely heritable (.61), with non-shared environmental factors having a substantial impact (.43) and shared environmental factors have a little effect (.07; King, Katz, Thompson, &

Macnmara, 2019). In other words, approximately 60% of the variability in spatial visualization can be accounted for, statistically, by genetic differences between people (in this particular sample). Even more germane to the current study, however, is the extent to which relations between spatial visualization and numerical abilities are due to shared genetics. This question was recently addressed by Tosto et al. (2014) through a twin study (N=1,539 monozygotic twins; N=2,635 dizygotic twins). As expected, they found a strong relation between spatial visualization skills and mathematical abilities, including measures of numerical reasoning ($r > .40$). Moreover, they found that approximately 60% of this overlap was explained by common genetic effects, while 40% of the overlap was due to environmental experience (26% and 14% by shared and non-shared environments respectively). Taken together, these studies suggest genetics may help explain individual differences in spatial visualization skills as well as common variance between spatial and numerical relations.

The malleability of spatial visualization

Although biological factors may place certain constraints on one's range of spatial visualization abilities, it is also clear that spatial skills are highly malleable constructs (Uttal et al., 2013). Compared to other core cognitive capacities, including working memory, spatial abilities – most notably spatial visualization skills – appear to be highly subject to practice and training effects. Evidence for this comes from Utall et al. (2013) who performed a meta-analysis examining the overall effects of 217 spatial training studies over a 25-year period (1984-2009). The study concluded that spatial training is an effective means for improving spatial thinking in people of all ages and across a variety of training techniques (e.g., video games, in-class training, spatial task training). The average effect size was large, approaching half a standard deviation (0.47). In theory, an improvement of this magnitude would approximately double the number of individuals with the spatial skills typically associated with being an engineer (Uttal et al., 2013). Moreover, the results revealed evidence of equal transfer for near and intermediate transfer measures. That is, training on one particular task, such as mental rotation, was found to not only lead to improvements in that same type of task, but resulted in improvements in untrained spatial tasks, such a mental paper folding (e.g., see Wright et al., 2008; Chu & Kita, 2011). In terms of durability, similar gains were observed immediately after training, less than 1 week delay, or less than one month delay. Moving forward, it will be important to assess just how long training related gains persists. The results of this study are important but puzzling.

They are important as the implications are significant and far reaching, especially in considering the ways in which spatial training might help boost STEM-related performance (as suggested in the engineering example above). The results are puzzling in that the effects of training spatial abilities appears unlike the training of any other cognitive abilities. To our knowledge, only spatial training has been found to reliably yield intermediate transfer effects. Inquiries why this is and what makes spatial thinking an especially malleable cognitive construct are needed. This information may be useful in designing educational curricula and interventions.

Does spatial visualization training transfer to mathematics?

Given the evidence that spatial thinking is highly malleable, might spatial training be an effective means to improve numerical thinking? Indeed, the answers to this question have the potential to provide key insights into the four candidate mechanisms reviewed. Unfortunately, there is no conclusive answer to this question. To date, the evidence is mixed and appears to very much depend on the training approach. Moreover, there are only two studies which meet the criteria for a randomized controlled trial, both of which included small sample sizes. In Cheng and Mix's (2014) study, 58 children aged 6–8 years underwent either 40 min of spatial visualization training (mental rotation) or 40 min of an active control task (crossword puzzles). Compared to the control group, children in the spatial training group demonstrated gains on a measure of spatial visualization, but, most impressively, also demonstrated improvements on calculation problems of two types: standard calculation problems (e.g., 56 + 6 = __) and missing term problems (e.g., 5 + __ = 12). Gains were more pronounced on the missing terms problems. In line with the spatial modeling account, the authors suggested that the intervention may have been effective because it encouraged participants to more effectively model the problems (e.g., reorganize 5 + __ = 12 into the more familiar question format, __ = 12-5. A follow-up study by Hawes et al., 2015 failed to replicate these effects. In this study, 61 6- to 8-year-olds were assigned to either 6 weeks (3 h total) of computerized spatial visualization training program or 6 weeks of computerized literacy training (control). Compared to the control group, children who received spatial training demonstrated improvements on spatial visualization measures but demonstrated no evidence of gains on any of the mathematics measures, including miss-terming problems. These mixed findings and the small sample sizes used make it clear that much more research is needed before any conclusions can be made about whether spatial training generalizes to numerical and mathematical reasoning. Moreover, future studies of this sort should aim to more explicitly address the potential mechanism(s) that may or may not facilitate transfer. For example, to further test the possibility that the spatial modeling account might be at play (as the authors of both the training studies above suggest), it is imperative to capture evidence of this. This may be achieved through self-report strategy use or through having participants write or draw their solution strategies.

The two studies above represent the only randomized controlled studies; however, three other studies have examined the effects of spatial visualization training on mathematical performance through classroom-based “quasi-experimental” studies. Because random assignment did not occur at the level of the individual, the effects of these studies may have been more influenced by uncontrolled variables (e.g., different teachers). A major benefit of these studies, however, is that they were carried out by classroom teachers and may be considered more ecologically valid approaches to spatial training. In the first of these studies, Hawes, Moss, Caswell, Naqvi, & MacKinnon, 2017 worked with kindergarten to Grade 2 teachers to implement a 32-week spatial visualization training intervention as part of teachers' regular mathematics instruction (total ~ 45 h of spatial training). Compared to an active control group (n = 28), children in the spatial training classrooms (n = 39) demonstrated widespread improvement on a variety of spatial reasoning measures as well as gains on a symbolic number comparison task. However, as noted by the authors, many of the spatial visualization activities also incorporated aspects of numerical reasoning, which may have influenced the results. In fact, it is possible that the greatest potential for mathematical learning lies in the combination and integration of spatial and numerical instruction. However, such an approach limits the conclusions we can make about the unique contributions of spatial visualization in the learning of mathematics. In a somewhat similar study by Lowrie et al. (2017), the authors also found some evidence of transfer to mathematics following an intensive in-class spatial training program with 10- to 12-year-olds (N = 186; 20 h of training over 10 weeks). Children in the spatial training classrooms, but not the control classrooms, demonstrated improvements in spatial visualization as well as a comprehensive measure of mathematics. However, the mathematics measure included a combination of items related to both numerical concepts as well as geometrical concepts. Thus, it is possible that the gains were due to changes in geometrical reasoning, arguably closely related to or even dependent on spatial reasoning, and not numerical reasoning. Lastly, a recent study by Cornu et al. (2017) failed to find any transfer to mathematical reasoning. Compared to children in the control kindergarten classrooms (n = 57), children in the spatial training kindergarten classrooms (n = 68) demonstrated some gains on near transfer spatial measures, but showed no evidence of improvements on seven separate measures of mathematics (e.g., counting, number comparison, number naming, arithmetic).

In another quasi-experimental study, Cheung, Sung, and Lourenco (2019) examined the effects of an at-home spatial visualization intervention with 6- to 7-year-olds (N = 62). Compared to an active control group, children who participated in the at-home mental rotation training demonstrated near transfer gains in mental rotation ability and far transfer to arithmetic performance. Critically, such transfer could not be attributed to general cognitive improvement, as no gains were observed on measures of nonsymbolic comparison, verbal working memory, or language ability following training. Relevant to the current review, the authors speculate that far transfer may have been due to improvements in children's ability to mentally model arithmetical relations and/or ground numerical information along a mental number line.

Overall, the results of these “quasi-experimental” studies are difficult to interpret and at this point in time, few conclusions can be drawn. It is clear, however, that when improvements do occur in mathematics (and this was true in the Cheng and Mix study as well), the mechanism of transfer is not well understood. In fact, not one training study to date has systematically addressed the question of what might mediate the effects of spatial visualization training on numerical reasoning. Thus, moving forward, it will be critical to target the underlying agents of change. The four candidate mechanisms reviewed here provide a good place to start. For example, different predictions can be made depending on the different accounts reviewed. According to the *spatial representation of numbers account*, one might predict that spatial training is related to improvements in one's internal representation of numbers according to a more spatially precise mental number line. This refinement in one's “mental number line,” in turn, is predicted to facilitate greater numerical reasoning. Critically, in order to test this hypothesis, future training studies will need to include measures of spatial-numerical mappings (e.g., intentional number-line estimation tasks, automatic SNA tasks, including SNARC effects). Any gains in more general measures of numerical reasoning should theoretically be mediated by change on these measures. As mentioned above, one way of testing the *spatial modeling account* would be to gain insights into the strategies that participants use while engaging in the numerical tasks. What evidence is there that the spatial visualization training actually led to an improved ability to mentally model the problem at hand? For example, collecting process data of the sort used in Hegarty and Kozhevnikov's (1999) word problem studies could be used to demonstrate the extent to which spatial visualization training results in improved schematic representations of the problems. Evidence of this sort would lend support for the spatial modeling account. In terms of the *shared neural processing account*, researchers have yet to examine the neural correlates of spatial training. However, a rather straightforward prediction would be that training-induced changes in neural activity (or the underlying neuroanatomical structures) should be correlated with improvements in numerical reasoning. Lastly, according to the *working memory account*, changes in spatial visualization should more broadly be encapsulated by changes in working memory. Indeed, it is possible that spatial visualization training is akin to working memory training. Future training studies thus need to also include measures of working memory to provide evidence for or against this possibility.

To conclude, future spatial training studies should look to go beyond simply measuring the effects of spatial training on numerical reasoning. Instead, trainings studies should be designed in ways that provide insights into the theorized mechanisms at play. This approach is critical in revealing why and under what conditions spatial training might be effective for some individuals but not others.

Potential mediators and moderators

As hinted at, the link between spatial visualization and numerical reasoning is likely to vary from individual to individual. For ease of clarification, this paper has only hinted at some of the potential mediators and/or moderators of the space-number relations. For example, we have suggested that spatial visualization may share stronger relations with unfamiliar vs. familiar numerical question types. With practice and experience, the need to engage visualization processes may be reduced. According to this proposal, the space-math link may differ across individuals as a function of their experience and familiarity with the mathematical task in question. For example, a child who is first learning basic arithmetic may find it useful to model the solution, whereas a child fluent in basic arithmetic may have no need to pause, reflect, and model the problem and solution. This suggests the need to not only consider the mathematical content under investigation, but also the learner's familiarity with the mathematical content when examining mechanisms underlying the space-math link. Said differently, mathematical experience may moderate relations between space and maths. To our knowledge, this represents a major gap in the literature and represents a promising area of future study.

In discussing the working memory account it was suggested that working memory or executive functions might mediate the relations between spatial visualization and mathematics. While research to date suggest evidence against this notion, more research is needed to more fully test this possibility. Moreover, it is possible that general intelligence (g) might account for the relations between space and mathematics. In fact, according to one account, intelligence might best be operationalized as the ability to spatially manipulate mental models (Lohman, 1996). Given prior findings of close relations between spatial visualization and general IQ (more specifically, nonverbal IQ), future research efforts are needed to further disentangle associations between spatial visualization abilities, general intelligence, and mathematics. To date, only one study has investigated this triad of relations and the results demonstrated strong and unique relations between spatial visualization and general mathematical abilities, even after controlling for g; Hawes et al., 2019a (as well as visual-spatial working memory). An important question moving forward is whether spatial visualization is related to mathematical reasoning due specific shared processes (e.g., the need to engage in mental rotation) or is related through more general processes (e.g., deductive reasoning). In addition to cognitive factors, many other variables might moderate the relation between spatial visualization and numerical reasoning, including, age, sex, type of mathematics instruction received in school (spatial vs. non-spatial), types of numerical reasoning required, past experiences with spatial learning, use of spatial versus non-spatial strategies, and various socioemotional and affective factors including spatial/mathematics anxiety. In combination, these factors represent a kaleidoscope of possible interactions. At present, we have only scraped the surface in studying how these and other variables may moderate space-number relations.

Conclusion

This paper highlights the potential ways in which spatial visualization and numerical abilities may be related to one another. Research is now needed to further probe and test the validity of these various accounts – both in isolation from one another but also in combination. Ultimately, it seems likely that all four accounts may offer insights into the ways in which spatial visualization and number are linked. In moving forward, it will not be enough to loosely base a study on one of the mechanistic accounts provided. For example, several studies to date have hypothesized strong relations between spatial abilities and mathematics because of research demonstrating shared neural resources. Indeed, as reviewed in this paper, there is good evidence to suggest that this is the case. However, we must go well beyond this level of theorizing: Not only stating which mechanism(s) are believed to underlie shared relations, but most critically, stating the precise ways in which the mechanism works to give rise to the relationship in question. A metaphor of a car mechanic helps to make this point. It is of use to know that car's mobility depends on its motor. This knowledge might help isolate the potential source of the problem. However, if the mechanic does not understand how the motor works, he/she has little chance of fixing a broken motor and regaining mobility. When it comes to expanding our understanding of spatial-numerical relations it is not enough to identify potential mechanisms that link spatial and numerical thought. The time is right to begin understanding why and under what specific conditions the mechanisms work, or just as importantly, fail to work. By better illuminating the learning processes that link spatial visualization and numerical competencies, we may be afforded new insights into the uniquely human ability to learn, perform, and invent abstract mathematics. This information, in turn, may prove critical in the assessment and design of effective mathematics curricula and intervention moving forward.

References

Alloway, T. P., & Passolunghi, M. C. (2011). The relationship between working memory, IQ, and mathematical skills in children. *Learning and Individual Differences*, *21*(1), 133–137.

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General*, *117*, 288–318.

Anderson, M. L. (2010). Neural reuse: A fundamental organizational principle of the brain. *Behavioral and Brain Sciences*, *33*(4), 245-266.

Anderson, M. L. (2016) Précis of. *Behavioral and Brain Sciences 39*.

Antonietti, A. (1999). Can students predict when imagery will allow them to discover the problem solution?. *European Journal of Cognitive Psychology*, *11*(3), 407-428.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617-645.

Barth, H. C., & Paladino, A. M. (2011). The development of numerical estimation: Evidence against a representational shift. *Developmental Science*, *14*(1), 125-135.

Bisiach, E., & Luzatti, C. (1978). Unilateral neglect of representational space. *Cortex*, *14*, 129- 133.

Boonen, A. J., van der Schoot, M., van Wesel, F., de Vries, M. H., & Jolles, J. (2013). What underlies successful word problem solving? A path analysis in sixth grade students. *Contemporary Educational Psychology*, *38*(3), 271-279.

Boonen, A. J., van Wesel, F., Jolles, J., & van der Schoot, M. (2014). The role of visual representation type, spatial ability, and reading comprehension in word problem solving: An item-level analysis in elementary school children. *International Journal of Educational Research*, *68*, 15-26.

Calabria, M., & Rossetti, Y. (2005). Interference between number processing and line bisection: a methodology. *Neuropsychologia*, *43*(5), 779-783.

Carpenter, P. A., & Just, M. A. (1986). Spatial ability: An information processing approach to psychometrics. In R. J. Stenberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 3, pp. 221-252). Hillsdale, NJ: Erlbaum.

Cheng, Y. L., & Mix, K. S. (2014). Spatial training improves children's mathematics ability. *Journal of Cognition and Development*, *15*(1), 2-11.

Cheung, C. N., Sung, J. Y., & Lourenco, S. F. (2019). Does training mental rotation transfer to gains in mathematical competence? Assessment of an at-home visuospatial intervention. *Psychological Research*, 1-18.

Chu, M., & Kita, S. (2011). The nature of gestures' beneficial role in spatial problem solving. *Journal of Experimental Psychology: General*, *140*(1), 102.

Cipora, K., Hohol, M., Nuerk, H. C., Willmes, K., Brożek, B., Kucharzyk, B., & Nęcka, E. (2016). Professional mathematicians differ from controls in their spatial-numerical associations. *Psychological Research*, *80*(4), 710-726.

Cipora, K., Patro, K., & Nuerk, H. C. (2015). Are spatial-numerical associations a cornerstone for arithmetic learning? The lack of genuine correlations suggests no. *Mind, Brain, and Education*, *9*(4), 190-206.

Cornu, V., Schiltz, C., Pazouki, T., & Martin, R. (2017). Training early visuo-spatial abilities: A controlled classroom-based intervention study. *Applied Developmental Science*, 1-21.

Davis, B, the Spatial Reasoning Study Group (2015). *Spatial reasoning in the early years: Principles, assertions, and speculations*. New York, NY: Routledge.

Dehaene, S. (2011). *The number sense: How the mind creates mathematics*. New York, NY: Oxford University Press.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, *122*(3), 371-396.

Dehaene, S., & Cohen, L. (2007). Cultural recycling of cortical maps. *Neuron*, *56*(2), 384-398.

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology, 20*(3-6), 487-506.

Delgado, A. R., & Prieto, G. (2004). Cognitive mediators and sex-related differences in mathematics. *Intelligence*, *32*(1), 25–32.

DeStefano, D., & LeFevre, J. A. (2004). The role of working memory in mental arithmetic. *European Journal of Cognitive Psychology*, *16*(3), 353–386.

Fischer, M. H., & Fias, M. H. (2005). Spatial representation of numbers. *Handbook of Mathematical Cognition*, 43-54.

Fischer, M. H., Mills, R. A., & Shaki, S. (2010). How to cook a SNARC: Number placement in text rapidly changes spatial–numerical associations. *Brain and cognition*, *72*(3), 333-336.

Fischer, U., Moeller, K., Bientzle, M., Cress, U., & Nuerk, H. C. (2011). Sensori-motor spatial training of number magnitude representation. *Psychonomic Bulletin & Review*, *18*(1), 177-183.

Font, V., Godino, J. D., & Gallardo, J. (2013). The emergence of objects from mathematical practices. *Educational Studies in Mathematics*, *82*(1), 97-124.

Galton, F. (1880). Visualised numerals. *Nature*, *21*(533), 252-256.

Galton, F. (1881). Visualised numerals. T*he Journal of the Anthropological Institute of Great Britain and Ireland, 10*, 85-102.

Gerstmann, J. (1940). Syndrome of finger agnosia, disorientation for right and left, agraphia and acalculia: local diagnostic value. *Archives of Neurology & Psychiatry*, *44*(2), 398-408.

Gevers, W., Reynvoet, B., & Fias, W. (2003). The mental representation of ordinal sequences is spatially organized. *Cognition*, *87*(3), B87-B95.

Gevers, W., Reynvoet, B., & Fias, W. (2004). The mental representation of ordinal sequences is spatially organised: Evidence from days of the week. *Cortex*, *40*, 171-172.

Giaquinto, M. (2008). Visualizing in mathematics. *The Philosophy of Mathematical Practice*. New York, NY: Oxford University Press.

Greeno, J. G. (1991). Number sense as situated knowing in a conceptual domain. *Journal for Research in Mathematics Education*, *22*(3), 170-218.

Gunderson, E. A., Ramirez, G., Beilock, S. L., & Levine, S. C. (2012). The relation between spatial skill and early number knowledge: the role of the linear number line. *Developmental Psychology*, *48*(5), 1229-1241.

Hadamard, J. (1945). *The psychology of invention in the mathematical field*. Princeton, NJ: Princeton University Press

Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbacher, M. A. (2007). The science of sex differences in science and mathematics. *Psychological Science in the Public Interest*, *8*(1), 1-51.

Hawes, Z., Moss, J., Caswell, B., & Poliszczuk, D. (2015). Effects of mental rotation training on children's spatial and mathematics performance: A randomized controlled study. *Trends in Neuroscience and Education, 4*(3), 60-68.

Hawes, Z., Moss, J., Caswell, B., Naqvi, S., & MacKinnon, S. (2017). Enhancing children's spatial and numerical skills through a dynamic spatial approach to early geometry instruction: Effects of a 32-week intervention. *Cognition and Instruction, 35*(3), 236-264.

Hawes, Z., Moss, J., Caswell, B., Seo, J., & Ansari, D. (2019a). Relations between numerical, spatial, and executive function skills and mathematics achievement: A latent-variable approach. *Cognitive Psychology, 109,* 68-90.

Hawes, Z., Sokolowski, H. M., Ononye, C. B., & Ansari, D. (2019b). Neural underpinnings of numerical and spatial cognition: An fMRI meta-analysis of brain regions associated with symbolic number, arithmetic, and mental rotation. *Neuroscience & Biobehavioral Reviews, 103,* 316-336.

Hegarty, M., & Kozhevnikov, M. (1999). Types of visual-spatial representations and mathematical problem solving. *Journal of Educational Psychology*, *91*(4), 684–689

Hegarty, M., & Waller, D. (2005). Individual differences in spatial abilities. *The Cambridge Handbook of Visuospatial Thinking*, 121-169.

Holmes, G. (1918). Disturbances of visual orientation. *The British Journal of Ophthalmology*, *2*(9), 449-468.

Hubbard, E. M., M. Piazza, P. Pinel, & Dehaene (2009). Numerical and spatial intuitions: a role for posterior parietal cortex? In L. Tommasi, L. Nadel, and M. A. Peterson (Eds.), Cognitive biology: Evolutionary and developmental perspectives on mind, brain and behavior (pp. 221–246). Cambridge, MA: MIT Press

Hutchison, J. E., Lyons, I. M., & Ansari, D. (2019). More similar than different: Gender differences in children's basic numerical skills are the exception not the rule. *Child Development*, *90*(1), e66-e79.

Huttenlocher, J., Jordan, N. C., & Levine, S. C. (1994). A mental model for early arithmetic. *Journal of Experimental Psychology: General*, *123*(3), 284-296.

Johnson-Frey, S. H. (2004). The neural bases of complex tool use in humans. *Trends in Cognitive Sciences*, *8*(2), 71-78.

Jordan, K., Heinze, H. J., Lutz, K., Kanowski, M., & Jäncke, L. (2001). Cortical activations during the mental rotation of different visual objects. *NeuroImage*, *13*(1), 143-152.

Just, M. A., & Carpenter, P. A. (1985). Cognitive coordinate systems: accounts of mental rotation and individual differences in spatial ability. *Psychological Review*, *92*(2), 137.

Kadosh, R. C., Lammertyn, J., & Izard, V. (2008). Are numbers special? An overview of chronometric, neuroimaging, developmental and comparative studies of magnitude representation. *Progress in Neurobiology*, *84*(2), 132-147.

Kaufman, S. B. (2007). Sex differences in mental rotation and spatial visualization ability: Can they be accounted for by differences in working memory capacity? *Intelligence*, *35*, 211–223.

Kell, H. J., Lubinski, D., Benbow, C. P., & Steiger, J. H. (2013). Creativity and technical innovation: Spatial ability's unique role. *Psychological Science*, *24*(9), 1831-1836.

Kersey, A. J., Braham, E. J., Csumitta, K. D., Libertus, M. E., & Cantlon, J. F. (2018). No intrinsic gender differences in children's earliest numerical abilities. *NPJ Science of Learning*, *3*(1), 12.

King, M. J., Katz, D. P., Thompson, L. A., & Macnamara, B. N. (2019). Genetic and environmental influences on spatial reasoning: A meta-analysis of twin studies. *Intelligence*, *73*, 65-77.

Knops, A., Viarouge, A., & Dehaene, S. (2009). Dynamic representations underlying symbolic and nonsymbolic calculation: Evidence from the operational momentum effect. *Attention, Perception, & Psychophysics*, *71*(4), 803-821.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity. *Intelligence*, 14(4), 389–433.

Kyttälä, M., Aunio, P., Lehto, J. E., Van Luit, J., & Hautamaki, J. (2003). Visuospatial working memory and early numeracy. *Educational and Child Psychology*, *20*, 65–76.

Kyttälä, M., & Lehto, J. E. (2008). Some factors underlying mathematical performance: The role of visuospatial working memory and non-verbal intelligence. European *Journal of Psychology of Education, 23*(1), 77–94.

Lakoff, G., & Núñez, R. E. (2000). *Where mathematics comes from: How the embodied mind brings mathematics into being.* New York, NY: Basic Books.

LeFevre, J. A., Jimenez Lira, C., Sowinski, C., Cankaya, O., Kamawar, D., & Skwarchuk, S. L. (2013). Charting the role of the number line in mathematical development. *Frontiers in Psychology*, *4*, 641.

Li, Y., & Geary, D. C. (2017). Children's visuospatial memory predicts mathematics achievement through early adolescence. *PLoS One*, *12*(2), e0172046.

Link, T., Moeller, K., Huber, S., Fischer, U., & Nuerk, H. C. (2013). Walk the number line–An embodied training of numerical concepts. *Trends in Neuroscience and Education*, *2*(2), 74-84.

Loetscher, T., Bockisch, C., Nicholls, M. E. R., & Brugger, P. (2010). Eye position predicts what number you have in mind. *Current Biology*, *20*, R264–R265

Lohman, D. F. (1988). Spatial abilities as traits, processes, and knowledge. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 4, pp. 181-248). Hillsdale, NJL Erlbaum.

Lohman, D. F. (1996). Spatial ability and G. In I. Dennis & P. Tapsfield (Eds.), *Human abilities: Their nature and assessment* (pp. 97–116). Hillsdale, NJ: Lawrence Erlbaum.

Lourenco, S. F., Cheung, C. N., & Aulet, L. S. (2018). Is visuospatial reasoning related to early mathematical development? *A critical review. In Heterogeneity of Function in Numerical Cognition* (pp. 177-210). London, UK: Academic Press.

Lowrie, T., Logan, T., & Ramful, A. (2017). Visuospatial training improves elementary students' mathematics performance. *British Journal of Educational Psychology*, *87*(2), 170-186.

Marghetis, T., Núñez, R., & Bergen, B. K. (2014). Doing arithmetic by hand: Hand movements during exact arithmetic reveal systematic, dynamic spatial processing. *The Quarterly Journal of Experimental Psychology*, *67*(8), 1579-1596.

Mathieu, R., Epinat-Duclos, J., Sigovan, M., Breton, A., Cheylus, A., Fayol, M., … Prado, J. (2017). What's Behind a "+" Sign? Perceiving an Arithmetic Operator Recruits Brain Circuits for Spatial Orienting. *Cerebral Cortex*, *28*(5), 1673-1684.

Mayer, E., Martory, M. D., Pegna, A. J., Landis, T., Delavelle, J., & Annoni, J. M. (1999). A pure case of Gerstmann syndrome with a subangular lesion. *Brain*, *122*(6), 1107-1120.

McCrink, K., Dehaene, S., & Dehaene-Lambertz, G. (2007). Moving along the number line: Operational momentum in nonsymbolic arithmetic. *Perception & Psychophysics*, *69*(8), 1324-1333.

Mix, K. S., & Cheng, Y. L. (2012). The relation between space and math: developmental and educational implications. *Advances in Child Development and Behavior*, *42*, 197-243.

Mix, K. S., Levine, S. C., Cheng, Y. L., Young, C., Hambrick, D. Z., Ping, R., & Konstantopoulos, S. (2016). Separate but correlated: The latent structure of space and mathematics across development. *Journal of Experimental Psychology: General*, *145*(9), 1206-1227.

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General*, *130*(4), 621-640.

Moss, J., Bruce, C. D., Caswell, B., Flynn, T., & Hawes, Z. (2016). *Taking shape: Activities to develop geometric and spatial thinking*. Grades K-2. Pearson Canada Incorporated.

Newcombe, N. S. (2010). Picture this: Increasing math and science learning by improving spatial thinking. *American Educator*, *34*(2), 29-35.

Newcombe, N. S., Levine, S. C., & Mix, K. S. (2015). Thinking about quantity: The intertwined development of spatial and numerical cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *6*(6), 491-505.

Paivio, A. (1983). The mind's eye in arts and science. *Poetics*, *12*(1), 1-18.

Paivio, A. (2013). *Imagery and verbal processes*. New York, NY: Psychology Press.

Patro, K., Fischer, U., Nuerk, H. C., & Cress, U. (2016). How to rapidly construct a spatial–numerical representation in preliterate children (at least temporarily). *Developmental Science*, *19*(1), 126-144.

Ramani, G. B., & Siegler, R. S. (2008). Promoting broad and stable improvements in low-income children's numerical knowledge through playing number board games. *Child Development*, *79*(2), 375-394.

Redick, T. S., Shipstead, Z., Wiemers, E. A., Melby-Lervåg, M., & Hulme, C. (2015). What's working in working memory training? An educational perspective. *Educational Psychology Review*, *27*(4), 617-633.

Root-Bernstein, R. S. (1985). Visual thinking: The art of imagining reality. *Transactions of the American Philosophical Society*, *75*(6), 50-67.

Schneider, M., Beeres, K., Coban, L., Merz, S., Susan Schmidt, S., Stricker, J., & De Smedt, B. (2017). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: A meta-analysis. *Developmental Science*, *20*(3), e12372.

Schneider, M., Merz, S., Stricker, J., De Smedt, B., Torbeyns, J., Verschaffel, L., & Luwel, K. (2018). Associations of number line estimation with mathematical competence: A meta-analysis. *Child Development*, 89(5), 1467-1485.

Seydell-Greenwald, A., Ferrara, K., Chambers, C. E., Newport, E. L., & Landau, B. (2017). Bilateral parietal activations for complex visual-spatial functions: Evidence from a visual-spatial construction task. *Neuropsychologia*, *106*, 194-206.

Shaki, S., Fischer, M. H., & Petrusic, W. M. (2009). Reading habits for both words and numbers contribute to the SNARC effect. *Psychonomic Bulletin & Review*, *16*(2), 328-331.

Siegler, R. S., & Ramani, G. B. (2009). Playing linear number board games—but not circular ones—improves low-income preschoolers' numerical understanding. *Journal of Educational Psychology*, *101*(3), 545-560.

Simms, V., Clayton, S., Cragg, L., Gilmore, C., & Johnson, S. (2016). Explaining the relationship between number line estimation and mathematical achievement: the role of visuomotor integration and visuospatial skills. *Journal of Experimental Child Psychology*, *145*, 22-33.

Sokolowski, H. M., Fias, W., Mousa, A., & Ansari, D. (2017). Common and distinct brain regions in both parietal and frontal cortex support symbolic and nonsymbolic number processing in humans: *A functional neuroimaging meta-analysis. Neuroimage, 146,* 376-394.

Stengel, E. (1944). Loss of spatial orientation, constructional apraxia and Gerstmann's syndrome. *Journal of Mental Science*, *90*(380), 753-760.

Tam, Y. P., Wong, T. T. Y., & Chan, W. W. L. (2019). The relation between spatial skills and mathematical abilities: The mediating role of mental number line representation. *Contemporary Educational Psychology*, *56*, 14-24.

Titze, C., Jansen, P., & Heil, M. (2010). Mental rotation performance and the effect of gender in fourth graders and adults. *European Journal of Developmental Psychology*, *7*(4), 432-444.

Tolar, T. D., Lederberg, A. R., & Fletcher, J. M. (2009). A structural model of algebra achievement: computational fluency and spatial visualisation as mediators of the effect of working memory on algebra achievement. *Educational Psychology*, *29*(2), 239–266.

Toomarian, E. Y., & Hubbard, E. M. (2018). On the genesis of spatial-numerical associations: Evolutionary and cultural factors co-construct the mental number line. *Neuroscience & Biobehavioral Reviews*.

Tosto, M. G., Hanscombe, K. B., Haworth, C. M., Davis, O. S., Petrill, S. A., Dale, P. S., … Kovas, Y. (2014). Why do spatial abilities predict mathematical performance?. *Developmental Science*, *17*(3), 462-470.

Uttal, D. H., & Cohen, C. A. (2012). Spatial thinking and STEM education: When, why and how. *Psychology of Learning and Motivation*, *57*, 147–181.

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013). The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin*, *139*(2), 352.

van Dijck, J. P., & Fias, W. (2011). A working memory account for spatial–numerical associations. *Cognition*, *119*(1), 114-119.

Viarouge, A., Hubbard, E. M., & McCandliss, B. D. (2014). The cognitive mechanisms of the SNARC effect: An individual differences approach. *PloS one*, *9*(4), e95756.

Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over fifty years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, *101*, 817–835.

Walsh, V. (2003). A theory of magnitude: common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences*, *7*(11), 483-488.

Wei, W., Yuan, H., Chen, C., & Zhou, X. (2012). Cognitive correlates of performance in advanced mathematics. *British Journal of Educational Psychology*, *82*(1), 157–181.

Wright, R., Thompson, W. L., Ganis, G., Newcombe, N. S., & Kosslyn, S. M. (2008). Training generalized spatial skills. *Psychonomic Bulletin & Review*, *15*(4), 763–771.

Zacks, J. M. (2008). Neuroimaging studies of mental rotation: a meta-analysis and review. *Journal of Cognitive Neuroscience*, *20*(1), 1-19.

Zorzi, M., Priftis, K., & Umiltà, C. (2002). Brain damage: neglect disrupts the mental number line. *Nature*, *417*(6885), 138-139.