Educational Research Review 20 (2017) 35–54

Contents lists available at ScienceDirect

Educational Research Review

journal homepage: www.elsevier.com/locate/edurev

Review

Homework and students' achievement in math and science: A 30-year meta-analysis, 1986–2015

Huiyong Fan [a], Jianzhong Xu [b], Zhihui Cai [c, *], Jinbo He [c], Xitao Fan [c]

[a] *Bohai University, China*
[b] *Mississippi State University, USA*
[c] *University of Macau, China*

A R T I C L E I N F O

Article history:
Received 2 August 2016
Received in revised form 31 October 2016
Accepted 4 November 2016
Available online 15 November 2016

Keywords:
Homework
Achievement
Math
Science
Meta-analysis
Synthesis

A B S T R A C T

In the current investigation, research conducted since 1986 were synthesized to examine the homework – achievement relationship in math/science, and to examine a range of factors that could have moderated this relationship. Our investigation revealed that there was an overall small and positive relationship between homework and academic achievement in math/science. Our investigation further revealed that the homework – achievement relationship in math/science was stronger for elementary and high school students than for middle school students. In addition, the homework – achievement relationship in math/science was shown to be the strongest in the studies involving US students, whereas it was the weakest in the studies involving Asian students. We discussed possible explanations for these and other findings, and the implications for future research directions.

Contents

* Corresponding author. Faculty of Education, University of Macau, Macao.
E-mail address: caizhihui922@163.com (Z. Cai).

http://dx.doi.org/10.1016/j.edurev.2016.11.003
1747-938X/

1. Introduction

Most typically defined as “tasks assigned to students by school teachers that are meant to be carried out during non-school hours” (Cooper, 1989, p. 7), homework is a common and widespread educational activity of everyday importance for many students, parents, and teachers across countries (Chen & Stevenson, 1989; Cooper, 1989; Dettmers et al., 2011; Núñez, Suárez, Cerezo, et al., 2015; Núñez, Suárez, Rosário, et al., 2015; Warton, 2001; Xu & Corno, 1998). It is thus not surprising that homework is frequently considered as an important instructional strategy to improve students' academic achievement (Cooper, Robinson, & Patall, 2006; Corno & Xu, 2004).

Yet, homework is a “complicated thing” (Corno, 1996), affected by more factors than any other instructional activities (Cooper, 2007). In addition, public attitudes toward homework have periodically shifted back and forth over the last century; news media and policy makers have a tendency to make generalized (and often polemical, anecdotal, and passionate) statements concerning homework (Gill & Schlossman, 2000, 2004; Warton, 2001). Not surprisingly, homework has become a perennial topic of public interest and debate, and an active area of investigation among researchers (Cooper, 1989; Cooper et al., 2006; Gill & Schlossman, 2004; Xu, 2012).

With two rare exceptions (i.e., Austin, 1979; Marshall, 1983), previous syntheses of homework research (e.g., Cooper, 1989; Cooper et al., 2006; Goldstein, 1960; Harding, 1979; Keith, 1986) have largely focused on homework in general (i.e., homework across different school subjects), without tapping into the influence of homework in specific subject matters (e.g., math and science). Yet, subject matter may influence homework's effect (Cooper, 1989; Goldstein, 1960; Harding, 1979; Paschal, Weinstein, & Walberg, 1984). Indeed, in their last synthesis concerning homework, Cooper et al. (2006) observed:

> “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.” (p. 28)

Concerning its limitations of generalizability, Cooper et al. (2006; p.53) added that their meta-analysis did not have sufficient number of studies for some subject areas, such as science. As a result, their conclusion about the small effect of subject matter on the homework – achievement relationship was tentative, and future studies on such subject matters would be needed.

Thus, it would be highly desirable to conduct a meta-analysis on the homework – achievement relationship in math/science for several additional reasons. First, some recent studies have called for and tapped into domain-specific aspects of homework (e.g., interest, effort, and self-regulation; Hong, Peng, & Rowell, 2009; Trautwein, Ludtke, Schnyder, & Niggli, 2006; Xu, Yuan, Xu, & Xu, 2014). Second, regardless of whether or not to pursue a postsecondary STEM field of study, math and science learning in K-12 is of critical importance; it has a powerful influence on curricular choices, educational and occupational opportunities, future employment, long-term career prospect, and expected future earnings (Chang, Singh, & Mo, 2007; OECD, 2012; Singh, Granville, & Dika, 2002). Third, math is an achievement domain with typically high homework demand (e.g., students usually spend about one-fifth to two fifths of homework time on math assignments; Kitsantas, Cheema, & Ware, 2011; Pezdek, Berry, & Renno, 2002; Xu, 2015). Finally, although there is still a need for homework research on science (Ramdass & Zimmerman, 2011), a substantial body of research has accumulated over the last decade that investigated the relationship between homework and math/science achievement (e.g., Areepattamannil & Kaur, 2013; Chang et al., 2007; Cheema & Sheridan, 2015; Trautwein, 2007; Wilson, 2010).

In the present investigation, we systematically examined previous research on the homework – achievement relationship in math/science. We then described our meta-analysis procedure, presented our analyses and the results of our meta-analysis. Finally, we discussed our results in the context of related literature, along with their implications for future investigation.

2. Literature review

We first consider the existing literature that links homework to math achievement and science achievement. Following that, we examine a range of potential moderator variables that may influence the relationship between homework and math/science achievement.

2.1. Studies linking homework and math achievement

Empirical findings concerning the relationship between homework and math achievement have been inconsistent. Many studies have found that homework is significantly and positively related to math achievement (e.g., Areepattamanil and Kaur, 2013; Austin, 1988; Cheema & Sheridan, 2015; Das, 2008; Drazen, 1992; Fernández-Alonso, Suárez-Álvarez, & Muñiz, 2015; Ladson, 2012; Mau & Lynn, 2000; Pelletier, 2005; Peng & Hill, 1995; Riley, 2007; Swank, 1999; Wong, 1993). For example, based on the data from National Longitudinal Study (NLS: 72), the High School and Beyond Study (HSB: 1980), the National Education Longitudinal Survey (NELS: 1988), Drazen (1992) found that time spent on homework was positively associated with math achievement.

In another study with 143 3rd graders, Pelletier (2005) investigated the relationship between homework performance (percent of homework completed and percent of homework correct) and math achievement (average test scores and standardized test scores), and found that homework performance was positively related to math achievement. Similarly, based on data from 7451 secondary school students in Spain, Fernández-Alonso et al. (2015) found that homework frequency, time spent on homework, and homework effort was positively related to math achievement.

Other studies, however, found that homework is either not related or negatively related to math achievement. For example, drawn from 1394 junior high school students from Netherlands, De Jong, Westerhof, and Creemers (2000) found homework time was negatively related to math achievement. Similarly, based on the U.S. sample from the Program for International Student Assessment (PISA 2003), Kitsantas et al. (2011) found time spent on homework was negatively associated with math achievement. Meanwhile, based on the Japan sample of TIMSS, the study by House (2002) revealed that homework frequency was not associated with math achievement, in line with similar findings from other studies (e.g., the linkage between homework completion and math achievement; Das, 2008; Swank, 1999).

Still other studies yielded contradictory findings (Olson, 1988; Omlin-Ruback, 2009; Rosário et al., 2015; Trautwein, 2007). For example, based on a sample of US students in grade 5, Omlin-Ruback (2009) examined the relationships between math achievement and homework (three types, including access skill practice as math homework, direct contact practice as math homework, and other homework). The study found that the correlation between math test and access skill practice was very low ($r = 0.01$), the correlation between math test and direct contact practice was moderate and positive ($r = 0.36$), and the correlation between math test and other homework was moderate but negative ($r = -0.28$). Likewise, Trautwein (2007) employed a nationally representative German sample in PISA program and reconsidered the relationship between homework and math achievement. The results indicated that the correlations were negative if homework was measured by time spent on homework, but the correlations were positive if homework was measured by students' effort.

2.2. Studies linking homework and science achievement

The relationship between homework and science achievement has been inconsistent as well. Some studies have found that homework is positively related to science achievement (Bertsos, 2005; Chang et al., 2007; Christensen, 2002; Fernández-Alonso et al., 2015; Mau & Lynn, 2000; ;Peng & Hill, 1995; Schibeci and Riley, 1986; Tamir, 1987; Walberg et al., 1986; Wilson, 2010). Using a random sample of 3135 17-year-olds from the 1976–1977 National Assessment of Educational Progress (NAEP) survey, Schibeci and Riley (1986) examined the correlation between time spent on homework and science achievement. Results revealed that homework was positively related to science achievement (0.20).

Recently, based on data from 7725 students in the second year of obligatory secondary education in Spain, Fernández-Alonso et al. (2015) examined the relationship between homework and science achievement. Homework was measured by frequency (from never to every day), time spent on homework (from less than 30 min to more than 120 min), and effort (using three Likert type items). Science achievement was measured through the use of a self-developed test. The study found that the correlation between homework and science test scores ranged from 0.07 to 0.18, and all correlations were positive and significant.

Meanwhile, other studies have found that homework is not associated with science achievement. Welch, Walberg, and Fraser (1986) used the data collected during 1981–1982 from a random sample of 1960 9-year-old students in a national assessment of educational progress in science. Results indicated that the correlation between the amount of homework (time spent on homework) and achievement was 0.03, $p > 0.05$. Similarly, based on data from 6031 13-year-old students from Hong Kong who were included in the TIMSS, House (2000) found the correlation between the frequency of homework and science scores was zero.

2.3. Potential factors related to inconsistent findings

A number of factors may have contributed to the inconsistent findings in previous research on homework and math/science achievement. These factors include: grade level, subject matter, homework indicators, publication type, publication year, sampling method, geographical region, and measure of achievement. The inclusion of these factors were informed by a temporal model of factors influencing the effects of homework (Cooper, 1989; Cooper et al., 2006), in which grade level and subject matter are exogenous factors that may influence the effectiveness of homework. Indicators of homework reflects assignment characteristics in Cooper's model. Measures of achievement corresponds to outcomes or effects in Cooper's model.

Publication type, publication year, and sampling method are important features of primary studies which are often included in meta-analysis (e.g., Card, 2011; Twenge, Zhang, & Im, 2004; see detailed discussion below). As studies conducted in different geographical regions are based on different student populations and are likely to moderate the homework-achievement relationship, the recent meta-analysis by Cooper et al. (2006) included only studies conducted in US. Consequently, it would be informative to include primary studies conducted in other geographical regions in our present meta-analysis.

2.3.1. Grade level

Previous syntheses of homework research suggest that the linkage between homework and achievement is moderated by grade level. For example, Cooper (1989) found that the average correlation between the amount of homework and student achievement as nearly $r = 0$ for elementary school students, $r = 0.07$ for middle school students, and $r = 0.25$ for high school students. Similarly, Cooper et al. (2006) found that the linkage between homework and achievement was stronger for students in Grade 7–12 ($r = 0.25$) than for students in K-6 ($r = -0.04$).

2.3.2. Subject matter

As discussed above, much of the previous syntheses of homework research (e.g., Cooper, 1989; Cooper et al., 2006; Harding, 1979; Keith, 1986) have not examined the homework – achievement relationship in specific subject matters. However, subject matter may play a role in homework's effect (e.g., Cooper et al., 2006; Paschal et al., 1984). This, to some extent, was substantiated by the findings that time and effort devoted to homework vary among different subject matters (e.g., Trautwein & Lüdtke, 2007; Trautwein, 2007).

2.3.3. Homework indicators

A range of indicators have been used in previous research on homework, including (a) homework time, (b) homework frequency, (c) homework completion, (d) homework effort, and (e) homework grade. Different studies may have used different indicators, which may have contributed to the inconsistent findings on the relationship between homework and achievement.

Time spent on homework. Under this operational definition, the amount of time (or number of hours) students spent on homework each week were recorded (Chang et al., 2007; Cheema & Sheridan, 2015; Mau & Lynn, 2000; Wong, 1993). For example, in the case of TIMSS 2007 data, the amount of time students spent on math homework was recorded from 1 (zero minutes) to 6 (more than 90 min) (Areepattamannil & Kaur, 2013).

Homework frequency. Under this definition, the frequency of homework that were assigned to the students each week was recorded. For example, homework frequency may range from “never” to “every day” (Areepattamannil & Kaur, 2013; Fernández-Alonso et al., 2015; House, 2000, 2002).

Homework completion. Under this definition, the percentage of homework that students completed was recorded (Austin, 1988; Rosário et al., 2015; Wilson, 2010). Sometimes teacher only considered the correct completion, not the total completion (e.g., Pelletier, 2005).

Homework effort. Homework effort was typically assessed using several items (e.g., “I make an effort to get good marks” and “I finish my homework even if they are difficult or they take me a long time”). Response options ranged from 1 (never or almost never) to 4 (always or almost always) (Fernández-Alonso et al., 2015; Trautwein, 2007).

Homework grades. Students' performance of homework was evaluated by their teachers. For example, teacher can give a score to a student's homework performance according to the completion or accuracy of homework (Bertsos, 2005; Christensen, 2002; Olson, 1988).

2.3.4. Publication type

Empirical studies on homework have appeared in different forms in a number of outlets, including journal article, conference paper or proceeding, dissertation, and technical report. Among these outlets, many journal articles, and to a less extent, conference papers, were peer-reviewed, whereas dissertations and technical reports were not. Given potential publication bias that peer-reviewed journals are more likely to publish articles with statistically significant findings (Card, 2011), publication type may potentially influence the observed effect sizes of homework on achievement in these different publication forms.

2.3.5. Publication year

For those research topics that are susceptible to societal influences, publication year is likely to a moderator, as the effect sizes of related empirical findings may change systematically over time. For example, based on 41 samples of children ages 9 to 14 and 97 samples of college students, Twenge et al. (2004) found that, along with publication year changed from 1960 to 2002, the young Americans increasingly believe their lives are controlled by outside forces (instead of their own efforts). In the case of homework, public attitudes toward homework have shifted back and forth over the last century (Cooper et al., 2006; Gill & Schlossman, 2004). For example, views of homework have shifted from a less positive assessment in the 1940s and in the mid-1960s (e.g., as an intrusion on other home activities and placing expressive pressure on children) to a more positive assessment in the late 1950s and in the mid-1980s (e.g., as a means of accelerating the pace of knowledge acquisition and helping students to complete in a global marketplace). Thus, publication year may potentially correlate with the effect sizes of homework on achievement.

2.3.6. Sampling method

Another likely factor for inconsistent findings concerning the relationship between homework and achievement could be sampling method, as sample representativeness may influence the accuracy of research findings (e.g., estimated standard errors) and the generalizability of these findings. Indeed, in the case of the present study, our preliminary review indicated that the correlations tend to be smaller for studies that used random samples (as compared with those that used non-random samples). Consequently, it would be important to incorporate sampling method as another potential moderator in the present meta-analysis.

2.3.7. Geographical region

Previous homework studies have involved samples from different geographical regions (e.g., US and Asia). Sampling region may become an important moderator in the present investigation, as there are regional differences in homework practices (Chen & Stevenson, 1989; Tam & Chan, 2009; Zhu, 2015). For example, based on data from TIMSS, Zhu (2015) found that the frequency of math homework varied among different regions. From 1995 to 2011, US teachers (compared with Japanese and Korean teachers) assigned math homework more frequently. On the other hand, US students (compared with Japanese and Korean students) spent much less time on homework (Walberg & Paschal, 1995) and performed less well on TIMSS assessments (Cooper & Schleser, 2006).

2.3.8. Measure of achievement

Achievement measures used in previous homework studies include both standardized tests (e.g., Austin, 1988; Cheema & Sheridan, 2015) and unstandardized assessments (e.g., Kitsantas et al., 2011; Wong, 1993). As the homework-achievement relationship may be stronger for unstandardized assessments than standardized tests (e.g., Cooper, Lindsay, Nye, & Greathouse, 1998), there is a need to incorporate measure of achievement as a potential moderator variable in the present study.

2.4. Study aims

In light of the inconsistent findings on the homework-achievement relationship in math and science areas published over the past several decades, the purpose of the current investigation is to quantitatively synthesize these studies for the purpose of providing an overall view on the homework-achievement relationship, and to study a range of factors that may have contributed to the observed inconsistencies by moderating this relationship. Specifically, the present investigation attempts to address the follow three questions. First, what is the overall correlation between homework and math/science achievement? Second, is this relationship identical across different grades or across different geographical regions? Third, is this relationship influenced by some other study features/characteristics such as publication year, publication type, sampling method, homework indicator, subject matter (i.e. math vs science), and measure of student achievement?

3. Method

3.1. Search for primary studies

In our search for the empirical studies that studied homework as related to math/science and students' achievement/performance in math/science, first, we used the digital databases of *ERIC*, *PsycINFO*, *ProQuest Dissertations & Theses Global*, *sciencedirect.com*, *link.springer.com*, *onlinelibrary.wiley.com*, and *tandfonline.com* to search for documents catalogued before December 7, 2015. The keywords "homework", "achievement", and "performance" were used both independently and jointly in the search. The initial search by using "homework" as the keyword returned about 8300 primary studies. Further search by using both "homework" and "achievement" (or "performance") narrowed the list to 2328 primary studies.

We conducted additional searches by using *Google Scholar* (scholar.google.com). We further used *Science Citation Index Expanded* and the *Social Sciences Citation Index* databases and traced and located all the research articles that cited the influential meta-analysis study on homework by Cooper (1989). In addition, we carefully compared the primary studies included in the meta-analysis study by Cooper et al. (2006) with those that were identified in our searches to make sure that our search did not miss any important primary studies on homework and achievement from the same time period. Based on our quick reading of the abstracts, and by using terms of "homework", "math (mathematics)", "science" (or "physics", "chemistry", "biology", "geography", "information", "technology") as word search in the text of the articles, we identified 115 primary studies that were potential candidates for inclusion into our meta-analysis. These 115 primary studies were set aside for later coding. Fig. 1 graphically describes our search process.

3.2. Inclusion and exclusion criteria

To be included in our meta-analysis, a primary study must satisfy the following criteria:

1. The participants of the study must be students in primary school, middle school, or high school (or equivalents). Studies that had college students as the participants were excluded. For example, Arasasingham, Martorell, and McIntire (2011) involved college students in their study on homework and science, and Babaali and Gonzalez (2015) examined the relationship between homework and grades in college pre-calculus class. These studies were not included in our study.
2. A study must have operationally defined homework and outcome measure(s). In addition, homework must have been clarified as homework related to math and science areas. If a study only used general homework (e.g., Cooper, Jackson, Nye, & Lindsay, 2001; Núñez, Suárez, Cerezo, et al., 2015; Núñez, Suárez, Rosário, et al., 2015; Xu, 2011) but without specifying that the homework was related to math/science, the study would not be included in our study. For example, Cooper et al. (2001) examined the relationship between homework completion and final grade, but it was about general homework, not homework related to math/science. As a result, it was excluded from our study.

If a study reported homework related to math/science, but if the homework variable was combined with, and cannot be separated from, other variables, then it was not included, because it would be impossible to isolate the effect of homework. For example, Das (2008) reported the relationship between the total time spent on both extra lessons and homework with

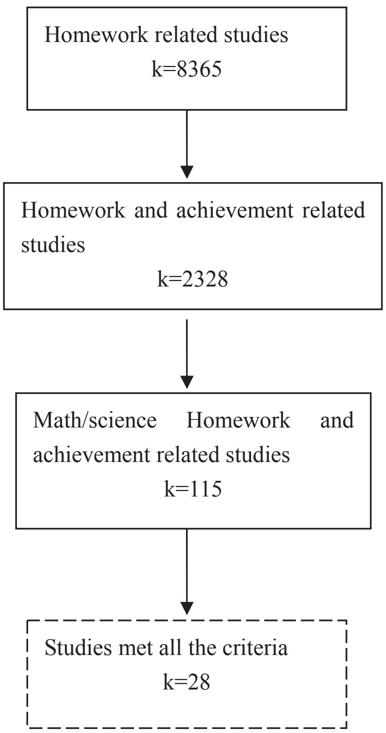

Fig. 1. Literature search process.

mathematics knowledge ($r = 0.343$, $p < 0.01$), not the relationship between homework and math knowledge. This result was not included in our meta-analysis, because we could not know the relationship of our interest (i.e., relationship between homework and math knowledge).

3. If a study reported more than one academic outcomes, and if these multiple outcomes were the repeated measurements of the same outcome at different time points (e.g., Trautwein, 2007), we only used the first measurement and its relationship with homework as the effect size to be included in our study. The reason for this decision was based on the consideration that the majority of other studies only reported the measure of academic outcome at one time point, and we would like to maintain the consistency with regard to this.
4. If a study reported the zero-order correlation between homework and academic outcome, or reported sufficient statistics to allow us to derive the zero-order correlation between the two, then the study would be included in the sample of primary studies for our meta-analysis. However, although some studies clearly defined and used appropriate homework and outcome measures, they only reported the results from regression modeling or multilevel modeling analysis (e.g., Maltes, Tai, & Fan, 2012), but did not provide sufficient statistical information for us to derive the zero-order correlation between homework and outcome. As a result, these studies could not be included into our sample of primary studies for the meta-analysis.
5. For inclusion into our meta-analysis sample, a study must have reported some basic but critical information. For example, sample size and grade (or age) of the study participants were essential for our purpose. Also, we only focused on the studies published in English; thus non-English publications were not searched or considered.

3.3. Information extracted from the primary studies

In order to understand the characteristics of the primary studies, and to search and identify the study features that could be potential moderators (Glass, 1976) for the inconsistent findings of the individual studies, we implemented a systematic process to identify and code the study features of each study included in our meta-analysis. The final set of study features that we considered are listed in Table 1. The process for coding these study features was described below.

3.4. Coding process and coding quality

Coding was done by two of the authors of the article. The process had three phases. The first is the "trial coding" phase. During this phase, the discussion of the research team, which included researchers in the substantive area of homework-

Table 1
Coded study features as potential moderators.

Publication Year
Continuous: 1986–2015
Publication type
Journal article
Dissertation
Conference presentation
Sampling method
Random
Non-random
Grade
Elementary
Junior high school
Senior high school
Mixed
Geographical region
USA
Europe
Asia
Subject
Math
Science
Indicators of homework
Homework completion
Frequency of homework
Homework effort
Homework grade
Time spent on homework
Others
Measures of achievement
Standardized test
Unstandardized assessment

related research and methodologists with experience in meta-analysis, a coding plan was tentatively worked out. The tentative coding table covered as many study features as the research team perceived as potentially relevant for later analysis. Based on this tentative coding table, an initial set of five primary studies were independently coded by the two coders. The research team then discussed about the issues encountered in the initial coding process, and made some needed revisions for the study features in the coding table. The second phase was to code all the studies based on the revised coding table, and the two coders independently coded all the primary studies on all the relevant study features initially. Later in this coding process, the two coders had exchanges and discussions about any issues encountered in the process. The third phase was for cross-checking/validation and integration of the coding results into the final coding table. During this phase, the coding results for each primary study were checked. Any issues noticed in the process would lead to the discussion of the coding team, and the relevant primary studies would be examined again for clarification and verification. During this phase, out of the 115 primary studies, five studies were shown to have varying levels of inconsistencies between the results of the two coders. This indicated that the coding consistency rate from the 2nd phase was 95.65% [i.e., (115-5)/115 = 95.65%], a very high degree of consistency between coders on multiple study features.

Based on the coding of the original 115 studies identified as relevant, 28 studies were finally identified to have met the inclusion criteria as described previously, thus useable for our meta-analysis. These 28 studies reported findings based on 41 independent samples, with the total accumulated sample size being 312,836. Five studies (Bertsos, 2005; Olson, 1988; Peng & Hill, 1995; Riley, 2007; Wilson, 2010) reported findings based on two independent samples, two studies (Chang et al., 2007; Drazen, 1992) reported findings based on three independent samples, and one study (Mau & Lynn, 2000) reported findings based on four independent samples.

It should be noted that some primary studies reported findings based on multiple indicators, with one effect size based on each indicator. More specifically, two primary studies (Areepattamanil and Kaur, 2013; Olson, 1988) reported two effect sizes corresponding to two different indicators. One primary study (Omlin-Ruback, 2009) reported three effect sizes for three indicators. One primary study (Trautwein, 2007) reported four effect sizes based on four indicators, and one study (Pelletier, 2005) reported five effect sizes based on five indicators. Finally, one primary study (Fernández-Alonso et al., 2015) reported eight effect sizes for eight different indicators. Altogether, with forty-one independent samples from the 28 primary studies, and with some containing multiple indicators from the same sample, the twenty-eight primary studies included in this meta-analysis produced a total sixty-one effect sizes representing the relationship between homework and academic performance in math or science. The appendix provides the detailed and complete coding table for all the studies included in this meta-analysis.

3.5. Effect size computation

In our quantitative synthesis, we intended to examine the relationship between homework in math/science and student performance in math/science. The ideal effect size measure is the zero-order correlation coefficient between homework measure and the measure of math/science performance. As described previously, some studies did not provide such correlation coefficient directly. For example, some studies used experimental design, and compared the performance under two conditions: one group of students with homework assignment, and a control group without homework assignment. A comparison between the two groups was made. In this situation, an effect size measure in the form of d[1] that is, standardized mean difference between the two groups, could be available. However, in such a case, it is necessary to convert d into r as shown below (Card, 2011), to be consistent with the effect size measures of other studies:

$$r = \frac{d}{\sqrt{4 + d}}$$

Once we obtained the correlation coefficient r between math/science homework variable and the variable for performance of math/science in a study, each r was weighted by using the weight w_i, which was based on the sample size associated with the r relative to the total accumulated sample size, as would be shown later in the Fisher Z transformation process. The overall weighted average correlation coefficient $\bar{r}$, and all subgroup average correlation coefficients in the moderator analyses, were obtained by going through several steps. First, each r_i went through the Fisher Z transformation:

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

The average Fisher $\bar{Z}$ was obtained as the weighted average Z_i:

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

[1] $d = \frac{M_1 - M_2}{S_{pooled}}$, where M_1 and M_2 are the means of Group 1 and 2, and S_{pooled} is the pooled standard deviation across the two groups.

Once $\overline{Z}$ was obtained, the average correlation coefficient between math/science homework and performance in math/science could be obtained through:

$$\bar{r} = \frac{e^{2Z} - 1}{e^{2Z} + 1}$$

Confidence intervals of $\bar{r}$ were obtained in similar fashion by using $SE_i = \frac{1}{\sqrt{w_i}}$ as the standard error for Z_i, and the lower and upper limits were obtained accordingly. The specialized statistical software, *Comprehensive Meta-Analysis* (CMA, Version 2.2), was used in the current study for conducting all the needed computations and analyses.

3.6. Heterogeneity test

The Q test was employed to evaluate the heterogeneity among effect sizes from the individual studies. Q test follows the χ^2 square distribution with degrees of freedom of $k - 1$ (Card, 2011; k: number of effect sizes used for the test). If the value of Q is larger than the critical value for the given degrees of freedom, it suggests that effect sizes are heterogeneous, and further moderator analysis for understanding what study feature(s) could account for such heterogeneity is needed. In our meta-analysis, we conducted the heterogeneity tests by using the built-in Q test function in CMA, which provided the needed statistical testing information.

3.7. Fixed-effects vs. random-effects models

Fixed-effects model and random-effects model are based on different assumptions. The fixed-effects model assumes that all the effect sizes from different samples/studies were from the same population with a common population effect size as its parameter. The differences among the sample effect sizes were the result of sampling error. Random-effects model, however, assumes that the effect sizes from different samples/studies could come from different populations, with each population having its own sampling distribution. As a result, there is no overall population parameter (i.e., no one overall population effect size), but there are different population parameters (i.e., multiple population effect sizes). As a result, the differences among the observed effect sizes were due to (1) true differences among multiple populations, and (2) sampling error. The applicability of each model depends on multiple considerations, such as the similarity of the samples used in the studies, the type of inferences that the researcher intends to make, etc. (Borenstein, Hedges, Higgins, & Rothstein, 2009; Card, 2011).

For our meta-analysis, we included studies about homework and achievement conducted across very different settings (e.g., cultural groups, grades, different subject areas), and it would be more reasonable to assume that the findings from these studies were not representing the same “population”. For example, previous research indicated that the tradition and value of homework in different cultures were different: parents in Chinese culture were more deeply involved in their children's homework than their counterparts in the US culture (e.g., Chen & Stevenson, 1989). As another example, in the US society, as part of the education reform in reaction to the Soviet Union's success in launching the first man-made satellite in the early 1950s, school homework received considerable emphasis in education (e.g., Pelletier, 2005), while other countries might not have such historical influence. Based on these considerations, it would be reasonable to consider the samples included in the present meta-analysis as coming from different populations. As a result, the random-effects model was adopted in the present meta-analysis.

3.8. Evaluation of publication bias

Publication bias refers to the potential issue that it would be easier for those studies with statistically significant findings to be accepted for publication, while those with statistically non-significant findings would have a harder time in getting accepted for publication; as a result, these studies would have higher probability of being left in the “file drawer”, thus the so called “file drawer problem” (Rosenthal, 1979). The present meta-analysis evaluated the potential publication bias of the collected studies included in the meta-analysis, as described in later sections.

3.9. Effect-size based meta-analysis and study-effect meta-analysis

As described above, our search yielded 41 independent samples from 28 studies, because a few studies contained more than one independent samples. In some cases, multiple indicators of math/science performance were used from one independent sample, resulting in more than one effect sizes (e.g., correlation coefficient) from an independent sample in some situations, leading to a total of 61 effect sizes. This means that those multiple effect sizes based on one study sample were not independent. We considered this situation, and conducted two types of meta-analysis. The first was based on all correlation coefficients between math/science homework and students' performance in math/science, and ignored the fact that some study samples had multiple effect sizes. For this meta-analysis, each independent study sample could contribute multiple correlation coefficients to the meta-analysis.

The second was a “study effects” meta-analysis in which, the multiple effect sizes based on the same study sample were averaged, and each study sample contributed only one effect size to the analysis. This approach is commonly known as "study-effects meta-analysis" (Bangert-Drowns, 1986). This approach was useful both for avoiding statistical dependence among the multiple effect sizes from the same study sample (Hunter & Schmidt, 1990), and for reducing the potential bias in favor of a few studies that had many effect sizes from the same sample. For our purpose, we used "independent sample" to define "multiple effect sizes", rather than a study itself. If a study contained two independent samples, and one effect size was produced from each independent sample, these two effect sizes were treated as independent, thus not averaged. Only when multiple effect sizes were produced from the same sample, these effect sizes would be averaged.

4. Results

Table 2 presents the major findings, including both effect-size based meta-analysis and study-effects meta-analysis. For the effect-size based meta-analysis ($k = 61$), the overall relationship between math/science homework and math/science performance was $r = 0.221$, with 95% confidence interval ranging from 0.191 to 0.250. The study-effects meta-analysis ($k = 41$) had very similar findings, with the overall relationship between homework and performance in math/science being 0.225, and the 95% confidence interval limits being 0.190 and 0.260. These findings suggested that, despite the inconsistencies across individual studies, overall, there was a positive, although somewhat weak, relationship between homework and performance in math/science areas.

The Q tests for heterogeneity among the effect sizes (both for the effect-size based analysis and for the study-effects analysis) were statistically significant, suggesting that there was considerable heterogeneity across the effect sizes, thus warranting the follow-up analysis for potential moderator variables that might have contributed to such inconsistencies.

Fig. 2 presents the "forest plot", a graphic description of the results based on the random-effects modeling analysis of the sixty-one effect sizes. In this forest plot, each effect size (square dot) and its estimated confidence interval (horizontal lines extending from both sides of the squared dot) was graphically shown. An effect size with short horizontal lines suggested a stable effect size with a narrow confidence interval, and those with long horizontal line indicated unstable effect sizes with wide confidence intervals. The "null" effect, that is, no relationship between homework and performance in math/science was represented by the vertical line of "0.00". The overall "effect" was represented by the dashed vertical line to the right of the "null" effect line, indicating an overall positive effect (i.e., a positive relationship between homework and performance in

Table 2
Relationship between homework and performance in math/science: Overall results and moderator analyses.

Moderators	Effect size based meta-analysis				Study-based meta-analysis			
	$Q_{between}$	k	Weighted $\bar{r}$	95% CI	$Q_{between}$	k	Weighted $\bar{r}$	95% CI
Overall	4166.86	61	0.221	(0.191, 0.250)	3387.79	41	0.225	(0.190, 0.260)
Publication type	633.38				514.06			
Journal article		30	0.139	(0.093, 0.185)		21	0.140	(0.079, 0.199)
Dissertation		24	0.432	(0.327, 0.526)		14	0.505	(0.377, 0.614)
Conference		7	0.205	(0.164, 0.245)		6	0.219	(0.180, 0.259)
Sampling method	242.08				197.03			
Random		19	0.168	(0.114, 0.220)		15	0.187	(0.131, 0.241)
Non-random		42	0.248	(0.212, 0.282)		26	0.258	(0.210, 0.304)
Grade	752.80				658.08			
Elementary		13	0.357	(0.183, 0.509)		6	0.321	(0.067, 0.535)
Junior high school		23	0.145	(0.106, 0.182)		13	0.149	(0.098, 0.198)
Senior high school		17	0.297	(0.253, 0.340)		17	0.297	(0.253, 0.340)
Mixed		8	0.074	(-0.155, 0.295)		5	0.067	(-0.212, 0.335)
Geographical region	1058.20				874.98			
USA		41	0.283	(0.248, 0.316)		31	0.271	(0.233, 0.307)
Europe		14	0.118	(0.067, 0.169)		5	0.074	(-0.015, 0.163)
Asia		6	0.075	(0.024, 0.126)		5	0.065	(0.011, 0.119)
Subjects	218.64				173.34			
Math		41	0.209	(0.170, 0.248)		24	0.196	(0.147, 0.245)
Science		20	0.233	(0.189, 0.276)		17	0.261	(0.209, 0.311)
Indicators of homework	471.04				796.85			
Homework completion		11	0.594	(0.361, 0.757)		5	0.691	(-0.029, 0.939)
Frequency of homework		5	0.117	(0.045, 0.188)		2	0.032	(-0.030, 0.094)
Homework effort		4	0.307	(0.225, 0.385)		/[a]	/	/
Homework grade		5	0.517	(0.417, 0.605)		4	0.525	(0.406, 0.626)
Time spent on homework		29	0.145	(0.107, 0.182)		21	0.185	(0.145, 0.225)
Others		7	0.072	(-0.155, 0.291)		9	0.156	(0.104, 0.207)
Measures of achievement	513.35				557.09			
Standardized assessment		40	0.196	(0.163, 0.227)		28	0.206	(0.172, 0.239)
Unstandardized assessment		21	0.297	(0.235, 0.357)		13	0.376	(0.263, 0.480)

Notes: k is the number of effect sizes involved. $\sum n_i$: Cumulative sample size of male across studies for each attitude type. All $Q_{between}$ tests were statistically significant at $\alpha = 0.001$. All Z tests for weighted correlations were statistically significant at $\alpha = 0.001$, except mixed grade subgroup and the subgroup of others under indicators of homework.

[a] All homework effort effect sizes were averaged as one of the others.

math/science). The diamond at the bottom of the overall "effect" line represents the width of the 95% confidence interval of the overall effect. As the left-side tip of the diamond did not cover the "null" effect line, the overall "effect" was statistically different from zero (i.e., statistically significant).

It was observed that about seven or eight effect sizes were negative (i.e., on the left side of the "null" effect line), and the remaining majority were positive. It was also obvious that some effect sizes had extremely wide confidence intervals (e.g., Olson, 1988; Swank, 1999), while others had very narrow confidence intervals (e.g., Areepattamanil and Kaur, 2013; Wilson, 2010). The width of a confidence interval is primarily a function of sample size of the study relative to the accumulated overall sample size across all the studies. The large difference in terms of the confidence interval widths suggested the necessity of weighting in meta-analysis, so that the potentially unstable effect sizes from small samples would not unduly influence the overall findings.

4.1. Assessment of publication bias

In our analysis, we used both a graphic approach (funnel plot) and Rosenthal's fail-safe *N* method to assess the potential publication bias. The funnel plot is a scatter plot of all the effect sizes relative to their standardized errors (Card, 2011). The magnitude of the standard errors was obviously a function of sample size. If there is no publication bias, the observed effect sizes (small circles in the plot) would distribute symmetrically around the vertical line; unsymmetrical distribution of the effect sizes would suggest the possibility of bias. The unsymmetrical pattern as shown in Fig. 3suggested potential bias in our data.

The trim-and-fill method is a two-step approach to correct potential publication bias (Duval & Tweedie, 2000). The first step is to draw the funnel plot as described above. This approach assumes that the most "negative" or "undesirable" studies were missing. As a result, the second step is to fill in some hypothetical effect sizes which were on the opposite side of the most positive effect sizes (the three solid dots on the left side of Fig. 2) until the funnel plot becomes symmetric. The results of the trim-and-fill method for our data indicated that, after filling in the hypothetical missing effect sizes (the three solid dots), the change of the overall average effect size was minimal (i.e., from 0.2126 to 0.2122), suggesting the potential publication bias in our data was negligible. A close look at the studies (Ladson, 2012; Wilson, 2010) with these three most positive effect sizes revealed that the samples underlying these large and positive effect sizes were very small (47–61). With the weighting of the effect sizes by sample size, these very large effect sizes actually had little influence on the estimation of the overall effect size.

The conclusion based on the funnel plot analysis was supported by the results of Rosenthal's fail-safe *N* method. The "fail-safe *N*" method estimates the number of unpublished studies with null results (i.e., zero effect sizes) that would be needed in order to bring the observed statistically significant overall effect size down to the level of being statistically non-significant. In our study, the Rosenthal's fail-safe *N* was calculated to be 76,742.6 (Fail safe $N = \left(\frac{\sum Z_i}{Z_\alpha}\right)^2 - k$, where Z_i is the standard normal deviate of significance from study *i*, Z_a is the threshold value of significance, *k* is the number of primary studies included.), which is much larger than the criteria that $5k+10 = 5 \times 61 + 10 = 315$ (Card, 2011, p. 270). In short, our assessment of the potential publication bias suggested that any publication bias would be minimal and negligible.

4.2. Moderator analyses

As shown above in Table 2, the overall heterogeneity test indicated that the effect sizes had considerable inconsistencies, which suggested the need for exploring potential moderators (i.e., study features) that could have contributed to such inconsistencies. As shown in Table 1, we considered and coded each study on a group of potential moderators.

4.2.1. Geographical region

As discussed previously, education practices in different countries or cultures could vary. For this reason, we considered the geographical region of a study sample. We grouped the effect sizes of the studies into three broad groups: US, Europe, and Asia. For both the effect-size based analysis and the study-based analysis, statistical heterogeneity among the effect sizes was shown ($Q_{between} = 1058.20$, $p < 0.01$, and $Q_{between} = 874.98$, $p < 0.01$ respectively for the two types of analysis). For both effect-size based and study-based analyses, of the three geographical regions, the studies with US samples had the largest effect size, while the studies with Asian samples showed the smallest effect sizes. In other words, the relationship between homework and math/science performance was shown to be the strongest in the studies involving US students (e.g., 0.283 for effect-size based analysis), while it was the weakest in the studies involving Asian students (e.g., 0.075 for effect-size based analysis).

4.2.2. Grade level

In homework research, age (or grade) of students is often an issues of research interest (Cooper, 1989; Cooper et al., 2006). In our study, we grouped the effect sizes based on the sample participants into three levels: elementary school students, middle school students, and high school students. A few studies had participants across these boundaries, and they were grouped into the "mixed" group. The heterogeneity test showed that the effect sizes varied more than just sampling variation ($Q_{between} = 752.80$, $p < 0.01$ and $Q_{between} = 658.08$, $p < 0.01$ for the two types of analyses, respectively). For the effect-size based analysis, the studies with elementary school students ($k = 13$) had the mean effect size of 0.357, vs. 0.145 for the middle school

Study name	Statistics for each study				
	Correlation	Lower limit	Upper limit	Z-Value	p-Value
Areepattamanil, 2013u1	0.150	0.122	0.178	10.246	0.000
Areepattamanil, 2013u2	0.080	0.051	0.109	5.435	0.000
Austin, 1988	0.368	0.076	0.602	2.442	0.015
Bertsos, 2005a	0.522	0.279	0.702	3.885	0.000
Bertsos, 2005b	0.458	0.197	0.659	3.282	0.001
Chang, 2007a	0.120	0.101	0.139	12.052	0.000
Chang, 2007b	0.110	0.089	0.131	10.198	0.000
Chang, 2007c	0.120	0.089	0.151	7.521	0.000
Cheema, 2015	0.340	0.315	0.364	24.975	0.000
Christensen, 2001	0.460	0.364	0.546	8.410	0.000
de Jong, 2000	-0.150	-0.208	-0.091	-4.969	0.000
Drazen, 1992a	0.195	0.181	0.209	27.225	0.000
Drazen, 1992b	0.294	0.287	0.301	72.956	0.000
Drazen, 1992e	0.196	0.184	0.208	31.395	0.000
Fernández-Alonso, 2015u1	0.130	0.108	0.152	11.489	0.000
Fernández-Alonso, 2015u2	0.260	0.239	0.281	23.384	0.000
Fernández-Alonso, 2015u3	0.200	0.178	0.221	17.815	0.000
Fernández-Alonso, 2015u4	0.180	0.158	0.201	15.992	0.000
Fernández-Alonso, 2015u5	0.070	0.048	0.092	6.161	0.000
Fernández-Alonso, 2015u6	0.180	0.158	0.201	15.992	0.000
Fernández-Alonso, 2015u7	0.170	0.148	0.192	15.085	0.000
Fernández-Alonso, 2015u8	0.140	0.118	0.162	12.384	0.000
House, 2000	0.000	-0.025	0.025	0.000	1.000
House, 2001	0.063	0.044	0.082	6.341	0.000
Hwang, 2011	0.540	0.236	0.748	3.253	0.001
Kitsantas, 2011	-0.170	-0.196	-0.143	-12.375	0.000
Ladson, 2012	0.679	0.515	0.795	6.300	0.000
Mau, 2000a	0.360	0.343	0.377	38.636	0.000
Mau, 2000b	0.280	0.262	0.298	28.907	0.000
Mau, 2000c	0.300	0.283	0.317	31.730	0.000
Mau, 2000d	0.230	0.211	0.248	23.532	0.000
Olson, 1988a	0.110	-0.032	0.248	1.514	0.130
Olson, 1988c	-0.080	-0.219	0.063	-1.099	0.272
Olson, 1988bu1	0.490	-0.008	0.793	1.933	0.053
Olson, 1988bu2	-0.140	-0.594	0.382	-0.508	0.611
Olson, 1988bu3	0.050	-0.457	0.533	0.180	0.857
Omlin-Ruback, 2009u1	0.010	-0.154	0.173	0.119	0.905
Omlin-Ruback, 2009u2	0.360	0.209	0.494	4.475	0.000
Omlin-Ruback, 2009u3	-0.280	-0.424	-0.122	-3.416	0.001
Pelletier, 2005u1	0.350	0.197	0.486	4.324	0.000
Pelletier, 2005u2	0.500	0.366	0.614	6.499	0.000
Pelletier, 2005u3	0.610	0.495	0.704	8.388	0.000
Pelletier, 2005u4	0.640	0.532	0.728	8.971	0.000
Pelletier, 2005u5	0.410	0.264	0.538	5.154	0.000
Pelletier, 2005u6	0.460	0.320	0.580	5.884	0.000
Peng, 1995a	0.230	0.221	0.239	47.747	0.000
Peng, 1995b	0.270	0.261	0.279	56.448	0.000
Riley, 2007a	0.220	0.181	0.259	10.700	0.000
Riley, 2007b	0.330	0.293	0.366	16.402	0.000
Schibeci, 1986	0.200	0.166	0.233	11.346	0.000
Swank, 1999	0.210	-0.244	0.588	0.904	0.366
Tamir, 1987	0.100	0.051	0.148	3.997	0.000
Trautwein, 2007u1	0.420	0.344	0.491	9.808	0.000
Trautwein, 2007u2	-0.240	-0.322	-0.154	-5.363	0.000
Trautwein, 2007u3	0.400	0.322	0.472	9.282	0.000
Trautwein, 2007u4	-0.300	-0.379	-0.217	-6.781	0.000
Walburg, 1986	0.190	0.147	0.232	8.498	0.000
Welch, 1986	0.030	-0.014	0.074	1.328	0.184
Wilson, 2010a	0.970	0.946	0.983	13.879	0.000
Wilson, 2010b	0.910	0.847	0.948	10.583	0.000
Wong, 1992	0.030	-0.017	0.077	1.260	0.208
Overall r	0.221	0.191	0.250	14.126	0.000

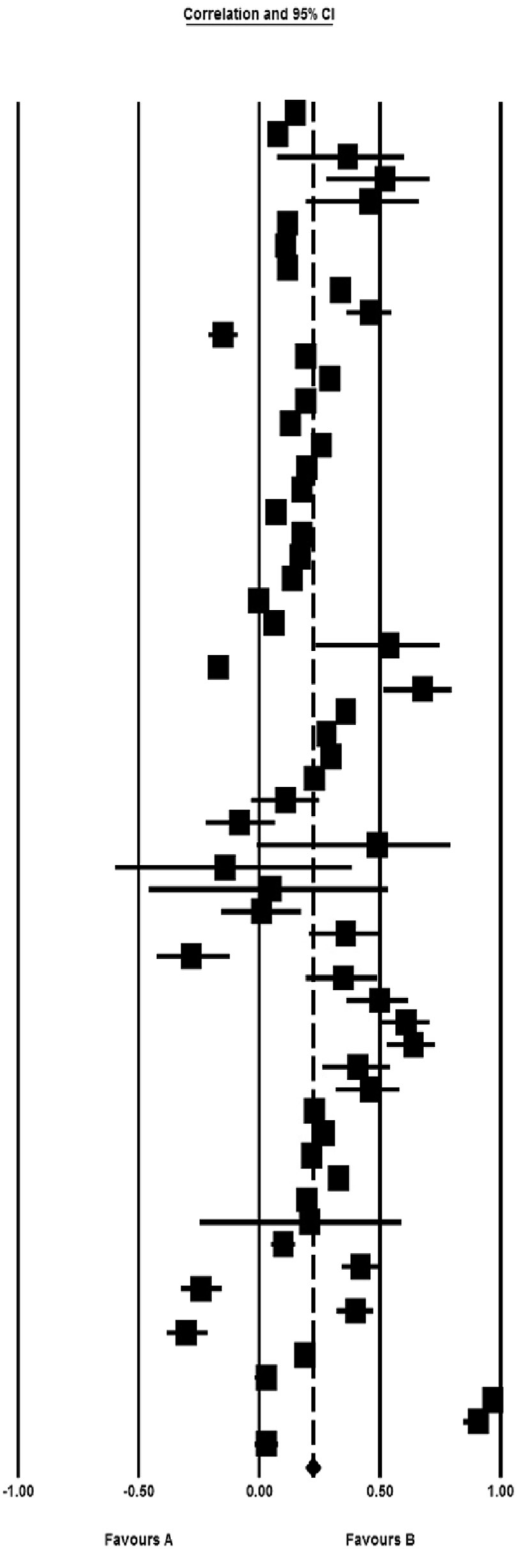

Fig. 2. Forest plot under random-effects model.

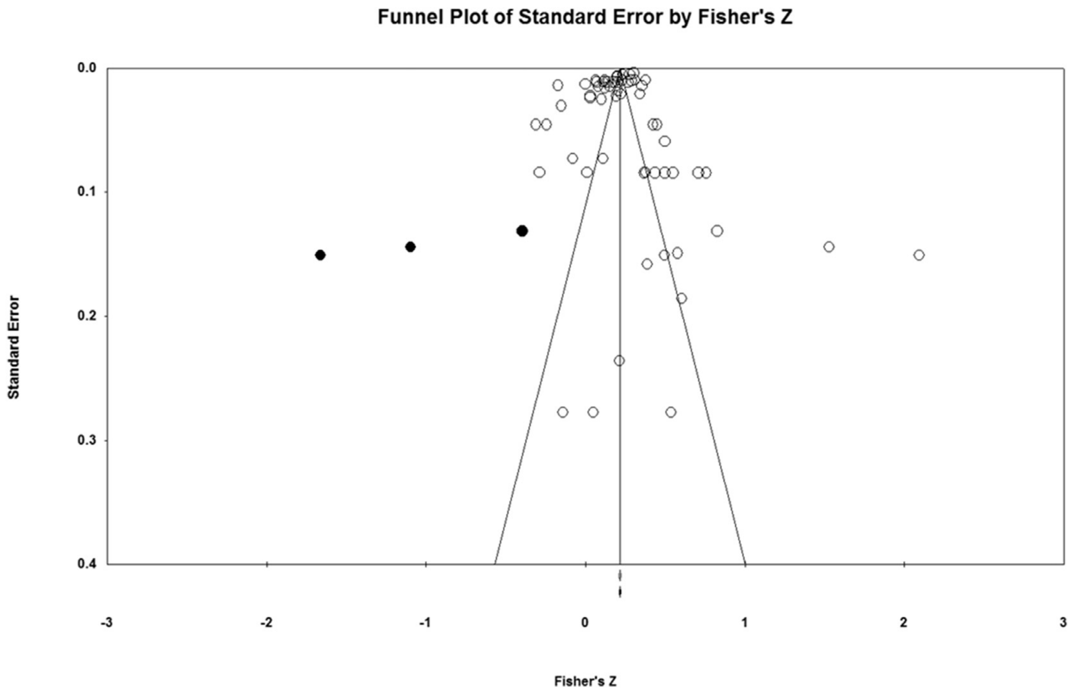

Fig. 3. Funnel plot of standard error by Fisher's Z.

group ($k = 23$), and 0.297 for the high school group ($k = 17$). The study-based analysis showed similar findings. The information from the confidence intervals (e.g., one condition's confidence interval did not include the point-estimate of another condition) suggested that the effect size from those studies involving middle school students was different from those involving elementary or high school students. More specifically, the results suggested that the relationship between homework and performance in math/science appeared to be stronger for elementary grades, while it was the weakest for middle school students.

4.2.3. Publication year

In a meta-analytic study covering a considerable time span, it is often of interest to see if the phenomenon of interest is associated with time. Our meta-analysis covered the studies for time frame of thirty years (1986–2015). It would be of interest to see if the relationship between math/science homework and performance in math/science was somehow related to time. Fig. 4 presents the scatter plot (and correlation coefficient) between publication years of the studies and effect sizes of the studies. The effect sizes fluctuated across years, and showed no identifiable trend (either increase or decline over the years) from 1986 to 2015. The correlation coefficient was 0.195 ($p = 0.132$) and statistically non-significant. The findings here suggested that the effect sizes of the studies did not appear to be related to time as represented by the publication year of a study.

4.2.4. Publication type

Of the forty-one studies collected for this meta-analysis, we had three types: journal articles (21), dissertations (14), and conference papers (6, including two technical reports). A total of 61 effect sizes were reported from these 41 studies, with 30 from the published journal articles, 24 from the dissertations, and 7 from the conference papers. The heterogeneity test for both effect-size based analysis ($Q_{between} = 633.38$, $p < 0.01$) and study-effects analysis ($Q_{between} = 514.06$, $p < 0.01$) suggested that the variation among the effect sizes as more than sampling error, and the levels (journal articles, dissertations, conference papers) of this moderator could have different effect sizes.

As shown in Table 2, for both effect-size based analysis and study-based analysis, the weighted average effect sizes from dissertations (=0.432 and 0.505 respectively) were the highest, and those from the journal articles (=0.139 and 0.140, respectively) were the lowest. The information from confidence intervals (e.g., one condition's confidence interval did not include the point estimate of another condition) suggested that the weighted average effect sizes of the three categories were statistically different from each other.

4.2.5. Sampling method

We considered random sampling vs. non-random sampling used in a study, and examined if the studies with these two sampling approaches had any differences in terms of their effect sizes (i.e., correlation between homework in math/science and math/science performance). Statistical heterogeneity among the effect sizes was shown both for the effect-size based

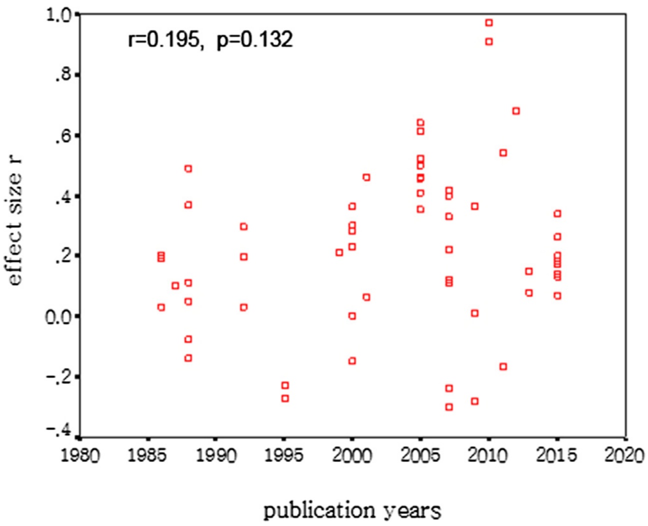

Fig. 4. Relationship between publication year and effect size.

analysis ($Q_{between} = 242.08$, $p < 0.01$ and) and for the study-based analysis ($Q_{between} = 197.03$, $p < 0.01$). As shown in Table 2, more studies used non-random sampling, and the weighted average effect size of those with non-random sampling was larger (=0.248 and 0.258 respectively for two types of meta-analyses) than those with random sampling (=0.168 and 0.187 respectively for two types of meta-analyses). Furthermore, the confidence intervals of the conditions (e.g., one condition's confidence interval did not include the point estimate of another condition) suggested that the differences between the two conditions were statistically significant.

4.2.6. Math vs. science

As we included empirical research articles that examined subject areas of either mathematics or science (including, e.g., physics, chemistry, biology, and information technology), we examined the issue whether there was any difference between these two subject areas. Although statistical heterogeneity among the effect sizes was revealed ($Q_{between} = 218.64$, $p < 0.01$, and $Q_{between} = 173.34$, $p < 0.01$, respectively for the two types of analysis), math and science areas showed relatively small difference (0.209 vs. 0.233 in effect-size based analysis, and 0.196 vs. 0.261 in study-based analysis). Confidence interval information suggested that the difference in the effect-size based analysis could be statistically non-significant (e.g., confidence interval of one condition covered the point estimate of another condition), but that the difference in the study-based analysis could be statistically significant (e.g., confidence interval of one condition did not cover the point estimate of another condition).

4.2.7. Homework indicator

For the "homework" variable, studies operationalized this variable differently. In our analysis, based on how homework was operationalized in a study, we grouped the studies into six groups, including homework completion, homework frequency, homework effort, homework grade, and time spent on homework, and "others" (those that could not be classified into any of the above), such as access skills, direct practice, distributed math homework, homework solutions, like homework, preparation for homework. The heterogeneity among the effect sizes was statistically confirmed. For the studies with different homework indicators, the analysis for this study feature indicated that the effect sizes from the studies with "homework completion," "homework grade," and "homework effort" were larger (e.g., 0.594, 0.517, and 0.307 in effect-sized based analysis), while those with "homework frequency" and "time spent on homework" had smaller effect sizes (e.g., 0.117 and 0.145 in effect-sized based analysis).

4.2.8. Measurement of achievement

As previous homework studies included both standardized tests and unstandardized tests, we examined whether there was any difference between these two achievement measures. Statistical heterogeneity among the effect sizes was shown for the effect-size based analysis ($Q_{between} = 513.35$, $p < 0.01$ and) and for the study-based analysis ($Q_{between} = 557.09$, $p < 0.01$). As indicated in Table 2, the weighted average effect size of those with standardized measure was larger (=0.297 and 0.376 respectively for two types of meta-analyses) than those with unstandardized measure (=0.196 and 0.206 respectively for two types of meta-analyses). In addition, the confidence intervals of the conditions (e.g., one condition's confidence interval did not include the point estimate of another condition) suggested that the differences between the two conditions were statistically significant.

5. Discussion

The current investigation has attempted to address an important gap in homework research, by synthesizing empirical studies on the homework – achievement relationship in two major subject matters (i.e., math and science) over the last three decades (i.e., 1986–2015). Our investigation revealed there was an overall small and positive relationship between homework and student achievement in math/science. This result is largely consistent with the previous findings on the homework – achievement relationship in general (Cooper, 1989; Cooper et al., 2006). On the other hand, as the synthesis by Cooper et al. (2006) was limited to studies conducted in US from 1987 to 2003 across different school subject areas, it can be argued that our investigation has expanded and extended Cooper et al.'s synthesis, in terms of time period (1986–2015 vs. 1987–2003), geographical region (US/Europe/Asia vs. US only), and subject matter (math/science vs. homework in general).

5.1. Explaining the grade level association

The grade level association found in the current investigation (i.e., the homework – achievement relationship in math/science was stronger for elementary and high school students than for middle school students) was intriguing. This finding is not in line with previous research syntheses finding that the homework – achievement relationship was stronger for high and middle school students than for elementary school students (Cooper, 1989; Cooper et al., 2006). Indeed, in these previous syntheses, the homework – achievement relationship was found to be negligible (i.e., nearly $r = 0$).

One possible explanation is that our study examined the homework – achievement relationship in math/science, whereas previous research syntheses tapped into the homework – achievement relationship in general (i.e., across different school subject matters). Math is among the school subject matters where teachers typically assign the largest amount of homework assignments (e.g., Pezdek et al., 2002; Xu, 2015) and has the strongest average correlation with homework variables when compared to other school subjects (Cooper, 2007). In addition, for younger students, more frequent and short homework assignments may be more effective than fewer but long assignments (Cooper, 1989, 2007); math homework tends to be short and more frequent in elementary grades. Thus, it seems logical that, for elementary school students, the homework – achievement relationship in math was stronger than that across different school subject matters.

Another related explanation concerns the purpose for math homework vs. homework in other subject matters in elementary grades. Compared with homework in other subject matters, elementary teachers are more likely to assign homework to improve immediate math achievement. This is to some extent substantiated by recent findings (Bedford, 2014), in which one 3rd-grade teacher, for example, commented that when she assigned math assignments on multiplication facts, "I'll just focus on one thing that's part of the common core [Common core State Standards for Mathematics]" (p. 88). Similarly, another 3rd-grade teacher stated his primary reason for assigning math homework:

> "A big reason for assigning homework is for students to practice the skills, especially the skills that we worked on in class or skills that I would like to move from basic understanding to foundational, so that students would know exactly what to do. In particular, something that we are now working on is ungrouping or regrouping." (p. 79)

How do we make out of the findings that the homework – achievement relationship in math was stronger for elementary school students than for middle school students? One likely explanation is related to parental involvement in homework at the elementary and middle school levels. Parents have greater mastery of the subject matter covered in the elementary grades (e.g., math and science; Cooper, 2001). Thus, in addition to two possible explanations discussed above, elementary students may benefit from parental involvement with homework (Patall, Cooper, & Robinson, 2008), whereas "middle school students generally may not benefit from parent involvement in homework" (Patall et al., 2008, p. 1089).

Furthermore, the transition to middle school is an especially difficult period for many students (Gutman & Midgley, 2000), who tend to experience a general decline in motivation, competence beliefs, and academic performance during the transition (Eccles et al., 1993). This is a more of case in the context of math learning. Declines in intrinsic motivation and competence beliefs in math are the steepest (Gottfried, Marcoulides, Gottfried, & Oliver, 2009) or particularly evident (Jacobs, Lanza, Osgood, Eccles, & Wigfield, 2002). Science is not immune to this decline, in the sense that "math and science are of particular concern because developmental decline in math and science motivation and attitudes has been a pervasive phenomenon across the literature" (Gottfried et al., 2009, p. 730). Thus, it is likely that this decline in math and science may influence the homework – achievement relationship in math/science at the middle school level.

It should be noted, however, that we need to be more cautious about interpreting the homework – achievement relationship in science for elementary school students, as we found only one study involving science achievement at the elementary school level (i.e., Welch et al.,1986). Clearly, there is a critical need to pursue a line of research that examines the linkage between homework and science achievement involving elementary school students.

5.2. Explaining the geographical region association

How do we explain the findings that the homework – achievement relationship in math/science was shown to be the strongest in the studies involving US students, whereas it was the weakest in the studies involving Asian students? These findings seem to be counterintuitive, given that Chinese students tend to hold more positive attitudes toward homework and spend more time on homework (Chen & Stevenson, 1989; Xu et al., 2014), and given that Chinese parents are more likely to encourage children to learn math, and to monitor and check their math homework (Cai, 2003).

Largely due to the perceived importance of family on education and the press on children's educational achievement, many Asian students have spent increasingly more time on after-school private tutoring (e.g., to supplement formal education and to prepare for high-stakes exams; Chen & Chang, 2015; Lee, 2007). For example, Koreans spent $6.8 billion on private tutoring in 1999, approximately a half of the government's education budget (Korean Ministry of Education, 2000; cited in Lee, 2007). In the case of Taiwan, more recently, Chen and Chang (2015) found that Taiwanese 5th graders spent about 70 min per day on homework, whereas they spent about 60 min per day on after-school private programs. This rapid increase in private tutoring has weakened the role of formal education in Asian countries (as compared with US). This is to some extent substantiated by the findings (Lee, 2007) that, based on the TIMSS survey, Korean 8th graders spent 3.6 times (72 min) as much time on private tutoring in math per week as did US 8th graders (20 min). Thus, it seems logical that the homework – achievement relationship is less pronounced for Asian students, particularly as private tutoring continues during the off-school periods.

5.3. Explaining the homework indicator association

With regard to homework indicators, our findings revealed that that the effect sizes from the studies with "homework completion," "homework grade," and "homework effort" were larger than those with "homework frequency" and "time spent on homework." These findings are largely consistent with previous homework literature (e.g., Cooper et al., 1998; Flunger et al., 2015; Trautwein, 2007; Xu, 2011, 2012).

The study by Cooper et al. (1998), for example, found that homework completion played a more important role in academic achievement than time spent on homework. Homework completion reported by secondary school students were positively related to course grades ($r = 0.31$, $p < 0.001$) and standardized test scores ($r = 0.14$, $p < 0.05$). As a comparison, time spent on homework was positively related to course grades ($r = 0.17$, $p < 0.01$), but was unrelated to standardized test scores ($r = 0.00$, $p > 0.05$). Similarly, at the secondary school level, Trautwein (2007) found that homework effort better predicted students' academic achievement than homework time, as weaker students tend to take longer to complete homework assignments, and as spending longer time on assignments may imply an inefficient or unmotivated working style.

As the combination of these homework indicators (e.g., "homework completion," "homework effort," and "time spent on homework.") has never been examined as a moderator in any previous homework meta-analysis, the unique contribution of the present meta-analysis lies in the examination and confirmation of these findings in the context of the homework – achievement relationship in math/science.

5.4. Other moderate variables

In addition to grade level and geographical region, the current investigation examined a number of study features/characteristics that may moderate the homework – achievement relationship in math/science. These moderator variables include publication year, publication type, sampling method, subject matter (math vs. science), homework indicators (i.e., homework completion, homework frequency, home effort, home grade, and time spent on homework), and measure of achievement. Our investigation revealed that, from 1986 to 2015, the effect sizes of the studies was not associated with a study's year of publication.

Concerning publication type, our analysis found that the weighted average effect sizes of the three categories (i.e., dissertation, conference presentation, and journal article) were statistically different from each other; those from dissertation were the highest, and those from the journal articles were the lowest. These findings here were contrary to the expectations of "publication bias" (i.e., published studies tend to have more statistically significant results, while non-published studies could have more statistically null results of no effect). However, based on our previous findings of lack of evidence for publication bias, the findings here were not surprising.

The analysis for the moderator variable of sampling method revealed that, in both effect-size based and study-based analyses, the studies that used random sampling approach had a smaller average effect size (i.e., lower average homework-achievement correlation coefficient) than those that used non-random sampling approach. If we believe that the studies with random sampling had more representative samples, this finding appears to suggest that the relationship between homework-achievement in math/science could be slightly lower than what the general literature would suggest. Because this issue (i.e., sampling method as a moderator variable) was not discussed in previous studies in homework research, this finding would be considered very tentative, and caution is warranted for interpreting this finding.

The comparison between the studies that respectively covered math or science as school subject matters indicated that the average effect size from the studies involving math is smaller than that from the others involving science. However, the difference is statistically non-significant (i.e., the estimate of one condition was within the confidence interval limits of the other condition) for the effect-size based analysis, but significant for the study-based analysis. Considering that effect-size based analysis could have data-dependency issue (i.e., multiple indicators from one sample), it is probably prudent to have more confidence in the finding of the study-based analysis.

6. Limitations and future research directions

The current investigation makes a valuable contribution to the rather fragmented body of empirical research literature on the homework – achievement relationship in math/science. On the other hand, some limitations should be noted. Although the current investigation included studies conducted in other countries (i.e., in addition to US), it was limited to studies written in English. The exclusion of publications in other languages may limit the generalizability of our findings to some extent. Second, in line with Cooper et al. (2006)'s observation a decade ago, a relatively small number of studies were found at the early elementary school grades, particularly concerning the homework – achievement relationship in science. Third, with respect to gender, we found too few studies that contained homework-achievement correlations separately for males and females. Thus, it is unclear whether the homework – achievement relationship in math/science is moderated by gender. Fourth, ideally, more detailed analyses (e.g., moderator analysis) could be done for different homework indicators (e.g., homework completion, frequency of homework,… time spent on homework, etc.) in order to develop better understanding about the relationship between homework and achievement in math and science. Unfortunately, the number of studies included in this meta-analysis would not allow us to do such more detailed analysis separately for each homework indicator, because the number of effect sizes would become too small for different levels of a moderator variable.

Consequently, it would be beneficial to conduct relevant studies concerning the homework and science achievement at the elementary school level, particularly given its recent emphasis on science learning at this developmental stage (Penfield & Lee, 2010; Xu, Coats, & Davidson, 2012). Meanwhile, as females are usually viewed to display less positive attitudes and interest in the stereotypical "male-dominated" domains such as math and science (Cleary & Chen, 2009), as few previous homework studies have examined the role of gender in the homework – achievement relationship in math/science, it would be beneficial to address this important gap in homework research.

Furthermore, as primary studies in general (with large scale survey studies in particular) do not provide detailed information concerning the type/complexity of homework assignments (e.g., long-term project, quantitative problems, and difficult level), we are unable to incorporate these important characteristics as moderators in our present study. Thus, it would be informative to incorporate these characteristics in future research on homework, thereby examining the likely interactions among certain important variables (e.g., whether difficult/challenging homework assignments have a greater effect for high school students in certain geographical regions). Finally, it would be highly desirable to incorporate qualitative studies that provide rich and detailed longitudinal descriptions of what happens concerning the homework – achievement relationship in math/science, as students move from elementary to middle school, and middle to high school, and in different geographical regions.

Acknowledgment

The research was supported by a grant (CPG grant RDG002/FED-FXT/2012) from the University of Macau to the last author, Xitao FAN. The first author was also supported partially by the National Natural Science Foundation of China (Grant No. 31100760). The authors of this paper deeply appreciate the supports.

Appendix

Characteristics of Primary Studies Included in the Meta-analysis

Study ID (1)	Publication Type (2)	Valid N	Female %	Age	Grade	Location	Sampling type	Subjects	Homework indicator	Achievement measure (3)	r
Areepattamannil & Kaur, 2013u1	C	4599	0.49		8	Singapore	Random	Math	Frequency	S	0.15
Areepattamannil & Kaur, 2013u2	C	4599	0.49		8	Singapore	Random	Math	Time	S	0.08
Austin, 1988	D	43			1	USA	Non	Math	Completion	S	0.37
Bertsos, 2005	D	48	0.48	15–16	10	USA	Non	Biology	Grade	UNS	0.52
Bertsos, 2005	D	47	0.6	15–16	10	USA	Non	Biology	Grade	UNS	0.46
Chang et al., 2007	J	9993	0.54		8	USA	Non	Science	Time	S	0.12
Chang et al., 2007	J	8529	0.54		10	USA	Non	Science	Time	S	0.11
Chang et al., 2007	J	3893	0.5		12	USA	Non	Science	Time	S	0.12
Cheema & Sheridan, 2015	J	4978	0.49		8–12	USA	Random	Math	Time	S	0.34
Christensen, 2002	D	289			7	USA	Non	Science	Grade	UNS	0.46
De Jong et al., 2000	J	1084		12.3	7	Netherlands	Non	Math	Time	UNS	−0.15
Drazen, 1992	C	19,000			12	USA	Random	Math	Time	S	0.2
Drazen, 1992	C	58,000			10	USA	Random	Math	Time	S	0.29
Drazen, 1992	C	25,000			8	USA	Random	Math	Time	S	0.2
Fernández-Alonso et al., 2015u1	J	7725	0.47	13.78 ± 0.82	8	Spain	Non	Math	Time	UNS	0.13
Fernández-Alonso et al., 2015u2	J	7725	0.47	13.78 ± 0.82	8	Spain	Non	Math	Frequency	UNS	0.26
Fernández-Alonso et al., 2015u3	J	7725	0.47	13.78 ± 0.82	8	Spain	Non	Math	Frequency	UNS	0.2
Fernández-Alonso et al., 2015u4	J	7725	0.47	13.78 ± 0.82	8	Spain	Non	Math	Time	UNS	0.18
Fernández-Alonso et al., 2015u5	J	7725	0.47	13.78 ± 0.82	8	Spain	Non	Science	Time	UNS	0.07
Fernández-Alonso et al., 2015u6	J	7725	0.47	13.78 ± 0.82	8	Spain	Non	Science	Frequency	UNS	0.18

(continued on next page)

(*continued*)

Study ID (1)	Publication Type (2)	Valid N	Female %	Age	Grade	Location	Sampling type	Subjects	Homework indicator	Achievement measure (3)	r
Fernández-Alonso et al., 2015u7	J	7725	0.47	13.78 ± 0.82	8	Spain	Non	Science	Effort	UNS	0.17
Fernández-Alonso et al., 2015u8	J	7725	0.47	13.78 ± 0.82	8	Spain	Non	Science	Time class	UNS	0.14
House, 2000	J	6031	0.45	13		Hong Kong	Non	Science	Frequency	S	0
House, 2002	J	10,107		13		Japan	Non	Math	Frequency	S	0.06
Hwang, Chen, Shadiev, & Li, 2011	J	32			7	Taiwan	Non	Math	Solutions	UNS	0.54
Kitsantas et al., 2011	J	5200	0.5	15.83 ± 0.29	7–11	USA	Random	Math	Relative time	UNS	−0.17
Ladson, 2012	D	61	0.48	10–11	5	USA	Non	Math	Grade	S	0.68
Mau & Lynn, 2000	J	10,512	1		10	USA	Random	Math	Time	S	0.36
Mau & Lynn, 2000	J	10,100	0		10	USA	Random	Math	Time	S	0.28
Mau & Lynn, 2000	J	10,512	1		10	USA	Random	Science	Time	S	0.3
Mau & Lynn, 2000	J	10,100	0		10	USA	Random	Science	Time	S	0.23
Olson, 1988u1	D	191			3–6	USA	Non	Math	Time	S	0.11
Olson, 1988u2	D	191			3–6	USA	Non	Math	Like	S	−0.08
Olson, 1988u1	D	16			3–6	USA	Non	Math	Grade	S	0.49
Olson, 1988u2	D	16			3–6	USA	Non	Math	Distributed homework	S	−0.14
Olson, 1988u3	D	16			3–6	USA	Non	Math	Preparatory homework	S	0.05
Omlin-Ruback, 2009u1	D	144			5	USA	Non	Math	Access skills	S	0.01
Omlin-Ruback, 2009u2	D	144			5	USA	Non	Math	Direct practice	S	0.36
Omlin-Ruback, 2009u3	D	144			5	USA	Non	Math	Other type	S	−0.28
Pelletier, 2005u1	D	143	0.44		3	USA	Non	Math	Completion	S	0.35
Pelletier, 2005u2	D	143	0.44		3	USA	Non	Math	Correct completion	S	0.5
Pelletier, 2005u3	D	143	0.44		3	USA	Non	Math	Completion	UNS	0.61
Pelletier, 2005u4	D	143	0.44		3	USA	Non	Math	Correct completion	UNS	0.64
Pelletier, 2005u5	D	143	0.44		3	USA	Non	Math	Completion	S	0.41
Pelletier, 2005u6	D	143	0.44		3	USA	Non	Math	Correct completion	S	0.46
Peng and Hill, 1995	T	41,571			8	USA	Non	Science	Time	S	0.23
Peng and Hill, 1995	T	41,571			8	USA	Non	Math	Time	S	0.27
Riley, 2007	D	2292	0		10	USA	Non	Math	Time	S	0.22
Riley, 2007	D	2292	0		12	USA	Non	Math	Time	S	0.33
Schibeici & Riley, 1986	J	3135		17	11	USA	Random	Science	Time	S	0.2
Swank, 1999 (4)	D	21	0.5	9–10	4	USA	Non	Math	Completion	UNS	0.21
Tamir, 1987	J	1590	56.5	10–17	11–12	Israel	Random	Science	Time	UNS	0.1
Trautwein, 2007u1	J	483	0.59	13.45	8	Germany	Random	Math	Effort	S	0.42
Trautwein, 2007u2	J	483			8	Germany	Random	Math	Time	S	−0.24
Trautwein, 2007u3	J	483			8	Germany	Random	Math	Effort	S	0.4
Trautwein, 2007u4	J	483			8	Germany	Random	Math	Time	S	−0.3
Walberg, Fraser, & Welch, 1986	J	1955	0.53	17		USA	Random	Science	Time	S	0.19
Welch et al., 1986	J	1960	0.5	9	3–4	USA	Random	Science	Completion	S	0.03
Wilson, 2010	D	47	0.47	15–17	11	USA	Non	Chemistry	Completion without graded	UNS	0.97
Wilson, 2010	D	51	0.55	15–17	11	USA	Non	Chemistry	Completion with graded	UNS	0.91
Wong, 1993	J	1766	0.44		7–13	Hong Kong	Non	Math	Time	UNS	0.03

Notes. (1) The u1~u8 at the end of study ID indicate that those studies report more than one outcome measures for each independent sample; while the a-d suggest that these studies include more than one independent sample. (2) D = Dissertation; J = Journal; T = Technical reports. (3) S=Standardized; UNS=Unstandardized; C=Conference. (4) Swank (1999) is from random experiment study, others from correlation study.

References[2]

Arasasingham, R. D., Martorell, I., & McIntire, T. M. (2011). Online homework and student achievement in a large enrollment introductory science course. *Journal of College Science Teaching, 40*(6), 70–80.

*Areepattamannil, S., & Kaur, B. (2013). Relationship of mathematics homework to mathematics achievement among grade 8 students in Singapore. In M. Inprasitha (Ed.), *Proceedings of the 6th East Asia regional conference on mathematics education* (pp. 363–370). Khon Kaen, Thailand: Khon Kaen University.

Austin, J. D. (1979). Homework research in mathematics. *School Science and Mathematics, 79*, 115–121.

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

*Austin, C. A. (1988). Homework as a parental involvement strategy to improve the achievement of first-grade children. *Dissertation Abstracts International, 50*(03A), 0622 (UMI No. 8911432).
Babaali, P., & Gonzalez, L. (2015). A quantitative analysis of the relationship between an online homework system and student achievement in pre-calculus. *International Journal of Mathematical Education in Science and Technology, 46*, 687–699.
Bangert-Drowns, R. L. (1986). Review of developments in meta-analytic method. *Psychological Bulletin, 99*(3), 388.
Bedford, P. D. (2014). *Teachers' beliefs and practices regarding homework: An examination of the cognitive domain embedded in third grade mathematics homework*. Unpublished Doctoral dissertation. University of Wisconsin-Milwaukee.
*Bertsos, G. F. (2005). Master's Theses and Doctoral Dissertations. *Differentiating biology homework to enhance academic achievement* (Vol. 126). Retrieved from: http://commons.emich.edu/cgi/viewcontent.cgi?article=1125&context=theses.
Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. West Sussex, United Kingdom: John Wiley & Sons.
Cai, J. (2003). Investigating parental roles in students' learning of mathematics from a cross-national perspective. *Mathematics Education Research Journal, 15*, 87–106.
Card, N. A. (2011). *Applied meta-analysis for social science research*. New York: Guilford Press.
*Chang, M., Singh, K., & Mo, Y. (2007). Science engagement and science achievement: Longitudinal models using NELS data. *Educational Research and Evaluation, 13*, 349–371.
*Cheema, J. R., & Sheridan, K. (2015). Time spent on homework, mathematics anxiety and mathematics achievement: Evidence from a US sample. *Issues in Educational Research, 25*, 246–259.
Chen, S. Y., & Chang, H. Y. (2015). Out-of-school time-use portfolios and Taiwanese children's reading achievement. *US-China Education Review, 5*, 336–348.
Chen, C., & Stevenson, H. W. (1989). Homework: A cross-cultural examination. *Child Development, 60*, 551–561.
*Christensen, E. F. (2002). *The effect of homework choices on achievement and intrinsic motivation* (Order No. 3030104). Available from ProQuest Dissertations and Theses Global. (250844898). Retrieved from: http://search.proquest.com/docview/250844898?accountid=12206.
Cleary, T. J., & Chen, P. P. (2009). Self-regulation, motivation, and math achievement in middle school: Variations across grade level and math context. *Journal of School Psychology, 47*, 291–314.
Cooper, H. (1989). *Homework*. White Plains, NY: Longman.
Cooper, H. (2001). Homework for all - in moderation. *Educational Leadership, 58*, 34–38.
Cooper, H. (2007). *The battle over homework: Common ground for administrators, teachers, and parents* (3rd ed.). Thousand Oaks, CA: Corwin.
Cooper, H., Jackson, K., Nye, B., & Lindsay, J. J. (2001). A model of homework's influence on the performance evaluations 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., Robinson, J. C., & Patall, E. A. (2006). Does homework improve academic achievement? A synthesis of research, 1987-2003. *Review of Educational Research, 76*, 1–62.
Cooper, J. A., & Schleser, R. (2006). Closing the achievement gap: Examining the role of cognitive developmental level in academic achievement. *Early Childhood Education Journal, 33*, 301–306.
Corno, L. (1996). Homework is a complicated thing. *Educational Researcher, 25*(8), 27–30.
Corno, L., & Xu, J. (2004). Doing homework as the job of childhood. *Theory into Practice, 43*, 227–233.
Das, I. (2008). *Exploration of the factors related to the cross-national mathematics achievement gap* (Order No. 3342774). Available from ProQuest Dissertations & Theses Global. (304449895). Retrieved from: http://search.proquest.com/docview/304449895?accountid=12206.
*De Jong, R., Westerhof, K. J., & Creemers, B. P. M. (2000). Homework and student math achievement in junior high schools. *Educational Research and Evaluation, 6*, 130–157.
Dettmers, S., Trautwein, U., Ludtke, O., Goetz, T., Frenzel, A. C., & Pekrun, R. (2011). Students' emotions during homework in mathematics: Testing a theoretical model of antecedents and achievement outcomes. *Contemporary Educational Psychology, 36*, 25–35.
*Drazen, S. (1992). *Student achievement and family and community poverty: Twenty years of education reform* (ERIC Document Reproduction Service No. 346234). Boston, MA: Annual Meeting of the Eastern Psychological Association.
Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*(2), 455–463.
Eccles, J. S., Midgley, C., Wigfield, A., Buchanan, C. M., Reuman, D., Flanagan, C., et al. (1993). Development during adolescence: The impact of stage-environment fit on young adolescents' experiences in schools and in families. *American Psychologist, 48*, 90–101.
*Fernández-Alonso, R., Suárez-Álvarez, J., & Muñiz, J. (2015). Adolescents' homework performance in mathematics and science: Personal factors and teaching practices. *Journal of Educational Psychology, 107*, 1075–1085.
Flunger, B., Trautwein, U., Nagengast, B., Lüdtke, O., Niggli, A., & Schnyder, I. (2015). The Janus-faced nature of time spent on homework: Using latent profile analyses to predict academic achievement over a school year. *Learning and Instruction, 39*, 97–106.
Gill, B. P., & Schlossman, S. L. (2000). The lost cause of homework reform. *American Journal of Education, 109*, 27–62.
Gill, B. P., & Schlossman, S. L. (2004). Villain or savior? The American discourse on homework, 1850-2003. *Theory into Practice, 43*, 174–181.
Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *The Educational Researcher, 10*, 3–8.
Goldstein, A. (1960). Does homework help? A review of research. *Elementary School Journal, 60*, 212–224.
Gottfried, A. E., Marcoulides, G. A., Gottfried, A. W., & Oliver, P. H. (2009). A latent curve model of parental motivational practices and developmental decline in math and science academic intrinsic motivation. *Journal of Educational Psychology, 101*, 729–739.
Gutman, L. M., & Midgley, C. (2000). The role of protective factors in supporting the academic achievement of poor African American students during the middle school transition. *Journal of Youth and Adolescence, 29*, 223–248.
Harding, R. C. (1979). *The relationship of teacher attitudes toward homework and the academic achievement of primary grade students*. Unpublished doctoral dissertation. Lehigh University.
Hong, E., Peng, Y., & Rowell, L. L. (2009). Homework self-regulation: Grade, gender, and achievement-level differences. *Learning and Individual Differences, 19*(2), 269–276.
*House, J. D. (2000). Relationships between instructional activities and science achievement of adolescent students in Hong Kong: Findings from the third international mathematics and science study (TIMSS). *International Journal of Instructional Media, 27*, 275–288.
*House, J. D. (2002). Relationships between instructional activities and science achievement of adolescent students in Japan: Findings from the third international mathematics and science study. *International Journal of Instructional Media, 29*, 275–288.
Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage Publications.
*Hwang, W. Y., Chen, N. S., Shadiev, R., & Li, J. S. (2011). Effects of reviewing annotations and homework solutions on math learning achievement. *British Journal of Educational Technology, 42*, 1016–1028.
Jacobs, J. E., Lanza, S., Osgood, W., Eccles, J. S., & Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. *Child Development, 73*, 509–527.
Keith, T. Z. (1986). *Homework*. West Lafayette, IN: Kappa Delta Pi.
*Kitsantas, A., Cheema, J., & Ware, H. W. (2011). Mathematics achievement: The role of homework and self-efficacy beliefs. *Journal of Advanced Academics, 22*, 310–339.
Korean Ministry of Education. (June, 2000). *Ka-yeol-kwa-woi ye-bang mit kong-kyo-yuk nae-silhwa bang-an jong-hap-bo-go-seo* [*Comprehensive report on plans for preventing excessive private tutoring and improving public education*]. Seoul, Korea: Author.

*Ladson, F. V. (2012). *The effect of homework completion on standardized math scores of fifth grade African American students* (Order No. 3499745). Available from ProQuest Dissertations & Theses Global. (954970135). Retrieved from: http://search.proquest.com/docview/954970135?accountid=12206.

Lee, J. (2007). Two worlds of private tutoring: The prevalence and causes of after-school mathematics tutoring in Korea and the United States. *Teachers College Record, 109*, 1207–1234.

Maltes, A., Tai, R. H., & Fan, X. (2012). When is homework worth the time? Revisiting the association between homework and achievement in high school science and math. *High School Journal, 96*(1), 52–72.

Marshall, P. M. (1983). *Homework and social facilitation theory in teaching elementary school mathematics*. Unpublished doctoral dissertation. Stanford University.

*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*(2), 119–125.

Núñez, J. C., Suárez, N., Cerezo, R., Gonzalez-Pienda, J. A., Rosario, P., Mourao, R., et al. (2015). Homework and academic achievement across Spanish compulsory education. *Educational Psychology, 35*, 726–746.

Núñez, J. C., Suárez, N., Rosário, P., Vallejo, G., Cerezo, R., & Valle, A. (2015). Teachers' feedback on homework, homework-related behaviors, and academic achievement. *Journal of Educational Research, 108*, 204–216.

OECD. (2012). *PISA 2012 Results in focus: What 15-year-olds know and what they can do with what they know*. Paris: OECD.

*Olson, M. E. (1988). *Homework variables and academic achievement: An integrated study* (Order No. 8821110). Available from ProQuest Dissertations & Theses Global. (303666289). Retrieved from: http://search.proquest.com/docview/303666289?accountid=12206.

*Omlin-Ruback, H. (2009). *A study of mathematics homework* (Order No. 3384674). Available from ProQuest Dissertations & Theses Global. (304979029). Retrieved from: http://search.proquest.com/docview/304979029?accountid=12206.

Paschal, R. A., Weinstein, T., & WAlberg, H. J. W. (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. (2008). Parent involvement in homework: A research synthesis. *Review of Educational Research, 78*, 1039–1101.

*Pelletier, R. (2005). *The predictive power of homework assignments on student achievement in grade three* (Order No. 3169466). Available from ProQuest Dissertations & Theses Global. (305350863). Retrieved from: http://search.proquest.com/docview/305350863?accountid=12206.

Penfield, R. D., & Lee, O. (2010). Test-based accountability: Potential benefits and pitfalls of science assessment with student diversity. *Journal of Research in Science Teaching, 47*, 6–24.

*Peng, S. S., & Hill, S. T. (1995). *Understanding racial-ethnic differences in secondary school science and mathematics achievement* (NCES Publication No. 95–710). Washington, DC: National Center for Education Statistics.

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.

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

*Riley, K. D. (2007). *An examination of time use patterns influence on academic achievement among African American and Hispanic male high school student athletes* (Order No. 3256940). Available from ProQuest Dissertations and Theses Global. (304743256). Retrieved from http://search.proquest.com/docview/304743256?accountid=12206.

Rosário, P., Núñez, J. C., Vallejo, G., Cunha, J., Nunes, T., Mourão, R., & Pinto, R. (2015). Does homework design matter? The role of homework's purpose in student mathematics achievement. *Contemporary Educational Psychology, 43*, 10–24.

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

*Schibeci, R. A., & Riley, J. P. (1986). Influence of students' background and perceptions on science attitudes and achievement. *Journal of Research in Science Teaching, 23*, 177–187.

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.

*Swank, A. L. G. (1999). *The effect of weekly math homework on fourth grade student math performance. Master of Arts action research project*. Knoxville, TN: Johnson Bible College. Retrieved from: http://files.eric.ed.gov/fulltext/ED433234.pdf.

Tam, V. C., & Chan, R. M. (2009). Parental involvement in primary children's homework in Hong Kong. *School Community Journal, 19*(2), 81.

*Tamir, P. (1987). Some factors which affect science achievement of high school seniors in Israel. *Research in Science & Technological Education, 5*(1), 69–92.

*Trautwein, U. (2007). The homework–achievement relation reconsidered: Differentiating homework time, homework frequency, and homework effort. *Learning and Instruction, 17*, 372–388.

Trautwein, U., & Lüdtke, O. (2007). Students' self-reported effort and time on homework in six school subjects: Between-student differences and within-student variation. *Journal of Educational Psychology, 99*, 432–444.

Trautwein, U., Lüdtke, O., Schnyder, I., & Niggli, A. (2006). Predicting homework effort: Support for a domain-specific, multilevel homework model. *Journal of Educational Psychology, 98*, 438–456.

Twenge, J. M., Zhang, L., & Im, C. (2004). It's beyond my control: A cross-temporal meta-analysis of increasing externality in locus of control, 1960-2002. *Personality and Social Psychology Review, 8*, 308–319.

*Walberg, H. J., Fraser, B. J., & Welch, W. W. (1986). A test of a model of educational productivity among senior high school students. *Journal of Educational Research, 79*, 133–139.

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

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

*Welch, W. W., Walberg, H. J., & Fraser, B. J. (1986). Predicting elementary science learning using national assessment data. *Journal of Research in Science Teaching, 23*, 699–706.

*Wilson, J. L. (2010). *The impact of teacher assigned but not graded compared to teacher assigned and graded chemistry homework on the formative and summative chemistry assessment scores of 11th-grade students with varying chemistry potential* (Order No. 3423989). Available from ProQuest Dissertations & Theses Global. (759967221). Retrieved from: http://search.proquest.com/docview/759967221?accountid=12206.

*Wong, N. Y. (1993). The relationship among mathematics achievement, affective variables and home background. *Mathematics Education Research Journal, 4*(3), 32–42.

Xu, J. (2011). Homework completion at the secondary school level: A multilevel analysis. *Journal of Educational Research, 104*, 171–182.

Xu, J. (2012). Homework and academic achievement. In J. Hattie, & E. M. Anderman (Eds.), *International guide to student achievement* (pp. 199–201). New York: Routledge.

Xu, J. (2015). Investigating factors that influence conventional distraction and tech-related distraction in math homework. *Computers & Education, 81*, 304–314.

Xu, J., Coats, L., & Davidson, M. L. (2012). Promoting student interest in science: The perspectives of exemplary African American teachers. *American Educational Research Journal, 49*, 124–154.

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

Xu, J., Yuan, R., Xu, B., & Xu, M. (2014). Modeling students' time management in math homework. *Learning and Individual Differences, 34*, 33–42.

Zhu, Y. (2015). Homework and mathematics learning: What can we learn from the TIMSS series studies in the last two decades? In J. A. Middleton, J. Cai, & S. Hwang (Eds.), *Large-scale studies in mathematics education* (pp. 209–234). Switzerland: Springer International Publishing.