Review of Educational Research
Month 201X, Vol. XX, No. X, pp. 1–37
DOI: 10.3102/0034654314540006

Effects of the Attributes of Educational Interventions on Students' Academic Performance: A Meta-Analysis

Hester de Boer
Anouk S. Donker
Margaretha P. C. van der Werf
University of Groningen

This meta-analysis examined the influence of attributes related to the implementation of learning strategy instruction interventions on students' academic performance, and also examined how the attributes related to the method of testing the intervention effects affected the actual effects measured. Using metaregression, we analyzed the influence of the subject domain in which the intervention was implemented, the implementer, its duration and intensity, student cooperation, and research method aspects (including measurement instrument). Most attributes moderated the intervention effect. Using forward regression analysis, we only needed four attributes to obtain the best model, however. This analysis showed that the intervention effect was lower when a standardized test was used for evaluation instead of an unstandardized test. Interventions implemented by assistants or researchers were more effective than those implemented by teachers or using computers. Cooperation had a negative, and session duration a positive, contribution. Together, these attributes explained 63.2% of the variance in effect, which stresses the importance of emphasizing not only the instructional focus of an intervention but also its other attributes.

Keywords: academic performance, meta-analysis, attributes of educational interventions and their effect studies, learning strategy instruction, moderator analysis

Large numbers of studies have been published in which the effects of educational interventions on student performance were tested. Meta-analyses have synthesized the results of these studies. The major emphasis of these meta-analyses is on the effect of the instructional focus of the interventions on student performance. However, other aspects can also influence the effectiveness of interventions. In the current meta-analysis, therefore, we also addressed the influence of attributes related to the way in which the instruction is offered and the study method. We examined the following: the subject domain in which the intervention was implemented, the implementer, the duration and intensity of the intervention, whether or not the students were allowed to cooperate, whether participants were randomly assigned to the experimental and control group, whether or not the fidelity of the implementation was checked, whether the control group perceived that they were taking part in an experiment or if it was "business as usual" for them, and, finally, the type of measurement instrument used to evaluate the effect.

Prior researchers may have addressed the influence of one or more of these attributes to some extent but not as thoroughly as we did. We included many attributes and analyzed their effects separately and in relationship to each other. Furthermore, we took important variables related to the instructional focus of the interventions into account as covariates. Although our meta-analysis only included learning strategy instruction interventions, the findings may be relevant to other types of educational interventions and their effect studies. Knowing which attributes of an educational intervention influence its effectiveness will help future researchers improve the quality of the implementation of their interventions. Knowledge about the effects of the more methodological attributes may contribute to the quality of future experimental designs in which intervention effects are estimated.

We start with an outline of prior research. Next, we formulate the research question and describe the research method. We then present the results and discuss our findings. Finally, we discuss the study's limitations and its practical implications and provide some suggestions for future research. This meta-analysis is an extension of the research by Donker, de Boer, Kostons, Dignath-van Ewijk, and van der Werf (2014) and is based on the same studies. Donker et al. (2014) dealt especially with the effects of the instructional focus of learning strategy instruction interventions on student performance. It showed which learning strategies are the most effective in enhancing student performance.

Prior Research

Learning strategies enable students to self-regulate their learning, and learning strategy interventions are aimed at improving student performance by enhancing students' self-regulated learning skills or metacognition. Three main categories of learning strategies can be distinguished. First, there are metacognitive learning strategies, such as the planning, monitoring, and evaluation of a learning task. Second, there are cognitive learning strategies, like rehearsal, elaboration, and organization. The third category encompasses management strategies that help students go through the entire learning process more successfully. For students to be able to use these learning strategies for self-regulated learning, metacognitive knowledge and motivation aspects are also important (Boekaerts, 1997; Mayer, 2008; Pressley, 2002; Weinstein & Mayer, 1986; Zimmerman, 2002). In addition to reporting on meta-analyses of these strategies, we also looked at other studies in which the effects of attributes of educational interventions were examined.

Subject Domain

The interventions analyzed in prior meta-analyses were often focused on different subject domains. For example, Chiu (1998) only selected studies on reading comprehension for his meta-analysis. Two other meta-analyses of learning strategy interventions by Hattie, Biggs, and Purdie (1996) and Dignath and Büttner (2008) did not select studies on the basis of a specific subject. The average effect sizes (Cohen's *d*) found in these analyses were 0.40 by Chiu, 0.57 by Hattie and colleagues, and 0.61 and 0.54 by Dignath and Büttner (primary school and secondary school, respectively). Dignath and Büttner also reported the average effect sizes for the subject domains separately: 0.44 and 0.92 for reading/writing in primary school and secondary school, respectively, and 0.96 and 0.23 for mathematics in primary school and secondary school, respectively. Dignath and Büttner's conflicting results for primary and secondary schools do not tell us much about differences in effectiveness related to subject domain. The lower effect size found in Chiu's study, however, might suggest that interventions are less effective in reading comprehension than in other subject domains, since the average effect sizes reported in the other studies are higher.

We found two meta-analyses that did not deal with learning strategy instruction, but which examined between-subject differences. Guskey and Pigott (1988) investigated the effect of mastery learning on student performance through a meta-analysis of 46 studies. Also comparing the effects for several subject domains, they found that there were significant differences among subjects. Interventions in the domain of mathematics yielded the highest effect (Hedges's $g = 0.70$), followed by language arts interventions (0.60), social studies (0.53), and science (0.50). The least effective were the psychology interventions (0.41). However, the within-subject domain differences were also quite large. Bangert-Drowns, Hurley, and Wilkinson (2004) conducted a meta-analysis of 48 school-based writing-to-learn interventions, examining whether the implementation and methodological characteristics of the interventions moderated the intervention effect on student performance. Their meta-analysis included studies conducted in mathematics, science, and social studies classes. The authors found no significant differences among subjects.

Implementer of the Intervention

Interventions also differ with respect to the provider of the instruction. Chiu (1998) reported that interventions had a higher effect size when they were implemented by the researcher rather than by the regular teacher (the difference in effect size was 0.24). Dignath and Büttner (2008) found a similar result for interventions implemented in secondary schools ($d = 0.80$) but observed no significant differences for interventions in primary schools. Bangert-Drowns et al. (2004), however, indicated that it did not matter whether or not the researcher was involved in the instruction.

Dynarski et al. (2007) summarized findings from prior research on computer-assisted instruction. These authors not only concluded that most studies showed a positive effect of using the computer as (supplemental) instructor but also noted that most of these studies were quite small or had methodological flaws. Subsequently, Dynarski et al. conducted a large-scale meta-analysis of the effect of computer programs on reading and mathematics in primary school. The meta-analysis tested the average effect of 16 software programs and included almost 12,000 participants. The results showed no significant effect of computer software compared with the traditional methods of the regular teachers.

The findings of some prior studies suggest that instructions may be more effective when provided by the researcher than by the teacher. The research literature gives several explanations for this possible relationship. First, teachers may be less motivated than researchers to test an intervention and, consequently, put less effort into it. The studies by Datnow and Castellano (2000) and Wehby, Maggin, Johnson, and Symons (2010) suggest that when teachers are offered a choice from among a number of interventions (e.g., a choice between two effective programs), or are able to co-construct the intervention, the implementation fidelity is higher, which leads to better results.

Abrami, Poulsen, and Chambers (2004) offer a related explanation. They examined the influence of several aspects of teacher motivation on implementation fidelity with respect to a cooperative learning program. They showed that the extent to which a teacher expected to be able to successfully implement an intervention had an influence on how well it was in fact done. Much less important was the perceived value of the intervention, referring to the degree to which a teacher perceives the intervention as worthwhile for the students or for himself or herself, a finding also reported by Datnow and Castellano (2000). Initial costs, like time and effort, did not play any role in how well teachers implemented a new program. Expectations regarding the intervention's success depended, among other factors, on how well the teachers were instructed by the researcher, and the amount of support they expected to receive during the implementation phase. As the researcher knows best how the intervention should be implemented, success expectancy might explain why interventions implemented by researchers yield higher effects.

A third possible explanation is that being taught by a person other than the regular teacher creates a novelty effect, which in turn has an influence on the students' performance, causing them to put more effort into their work. This novelty effect might result in a higher effect size for interventions implemented by researchers. We found a final explanation in Chiu's (1998) study: Researcher-implemented interventions might produce higher results because researchers are more inclined to teach to the test than teachers are.

Duration and Intensity of the Intervention

Hattie et al. (1996) found that the duration of the intervention matters. In their study, short programs (1 or 2 days) appeared to be more effective than interventions of 3 or 4 days, whereas even longer interventions (between 4 and 30 days) proved the most effective. Unfortunately, the authors did not present the results of the performance outcome measure separately. Instead, they combined performance, study skills, and affect. It is, therefore, not certain whether this finding also holds for performance only. In addition, after reviewing the effects of various types of programs in reading and mathematics for both primary and secondary school students, Slavin and Lake (2009) found a duration effect. Shorter programs (with a minimum duration of 12 weeks) had a greater effect than longer ones (with a duration of at least 1 year).

Dignath and Büttner (2008) observed that in both primary and secondary schools, the duration of interventions in mathematics had a small effect on the students' performance. Interventions including more sessions had a higher effect. However, intervention duration did not have a significant influence on the effect size with respect to reading/writing. Moreover, when the interventions for reading/writing and mathematics were merged into a single analysis, no significant effect of intervention duration was found.

Chiu (1998), Guskey and Pigott (1988), and Bangert-Drowns et al. (2004) found no significant effect for the duration of the intervention on its effectiveness. Chiu (1998), however, did find a small effect for intervention intensity, operationalized as the number of session days per week. Less intensive interventions were slightly more effective than more intensive programs. Bangert-Drowns et al. (2004) found no significant effect for intensity.

Cooperative Learning

After reviewing a large amount of research on cooperative learning, Slavin (1991) concluded that this approach is an effective method of teaching to enhance student performance. In learning strategy instruction interventions, it is quite common for cooperation to be used as a tool to support strategy learning. In addition, Dignath and Büttner (2008) argued that interventions based on cooperative learning produced a lower effect in primary school ($d = -0.24$) but a higher effect in secondary school ($d = 0.56$) compared with trajectories in which students were not stimulated to cooperate. A closer look at the data, however, showed that the negative effect in primary school was not significant.

Research Method

Hattie et al. (1996) examined whether the effect size of the interventions depended on the research design. They found that studies in which a control group or a pretest and posttest were used yielded a slightly lower effect size on the outcome measure than studies based on other designs. However, these differences were not significant. The outcome measure consisted of three aspects, namely, performance, study skills, and affect.

Chiu (1998) investigated if the effect size was related to whether or not the participants or classrooms in the primary studies were randomly assigned to the control group and the experimental group. When the random assignment was the only variable in the regression analysis, no significant relationship was found, but after correction for other training characteristics, it appeared that studies based on random assignment produced a lower effect size than those using other methods. Chiu also examined whether the impact of the intervention could be explained by the phenomenon of participants in a study improving or modifying their behavior because they know that they are being monitored. The coders were asked to indicate whether they thought the control group did or did not believe that it was receiving training. Unfortunately, the interrater reliability was too low to use this information in the meta-analysis.

Measurement Instrument

Hattie et al. (1996) reported a higher effect size when the training task and the performance goal were more closely related. Therefore, tests that measured near transfer yielded a higher effect size than those that measured further transfer of the learned task. In an analysis of 43 learning strategy instructions, Chiu (1998) found that the type of measurement instrument influenced the size of the intervention effect. Nonstandardized tests resulted in an average d of 0.61, whereas standardized tests produced an effect size of only 0.24. Moreover, this difference increased with correction for other intervention attributes.

Slavin and Madden (2011) observed a higher effect size for treatment-inherent measures than for treatment-independent measures. They reviewed the studies included in three meta-analyses by What Works Clearinghouse (WWC), of reading programs for beginners and elementary and middle school mathematics. Treatment-inherent measures focused on the content taught in the experimental group but not in the control group, whereas treatment-independent measures emphasized the content equally in the experimental and the control groups. The authors observed that treatment-inherent measures were most commonly used to test curricular changes. Experimenters who wished to test an instructional program, however, might be more inclined to use a treatment-independent test.

The Current Study

The studies described above show that the effectiveness of an intervention depends, at least to a certain extent, on attributes related to how the intervention is implemented and to the method of investigating its effect. However, not all studies yield the same picture. In the current meta-analysis, we systematically examined to what extent these attributes moderated the effects of interventions on student performance. Based on the results of this analysis, future intervention research can be optimized, and interventions can be implemented under optimum conditions, yielding a higher effect on students' academic performance. Our research question was as follows: Is the effect of an educational intervention on students' academic performance influenced by the attributes of the intervention and its effect study?

We examined the influence of two groups of attributes. The first group was made up of attributes related to the implementation of interventions. We distinguished the following attributes: (a) the subject domain in which the intervention was implemented, (b) the implementer, (c) the duration of the intervention, (d) the intensity of the intervention, and (e) whether or not students cooperated. Although this last attribute may seem a bit of an outlier, we included it because we noticed that many learning strategy instructions in our meta-analysis used cooperation as a supportive teaching method. The second group of attributes we examined related to the study method: how the effects of interventions were examined. We included the following attributes: (f) whether students were assigned randomly to the experimental and control groups, (g) whether the fidelity of the implementation was checked, (h) the perspective of the control group (business as usual or not), and (i) the type of measurement instrument used to evaluate the effect. We did not analyze differences related to the inclusion or exclusion of a control group or a pretest, as we only selected studies for our meta-analysis in which a control group and a pretest and posttest were used.

We used the data of a prior meta-analysis by Donker et al. (2014), who focused on the effect of learning strategy instruction on student academic performance. Donker et al. (2014) examined the effects of 14 learning strategies on student performance and analyzed whether the intervention effect was related to the characteristics of the students. The findings showed that learning strategy instruction that taught students how, when, and why to use learning strategies (i.e., general metacognitive knowledge), and to plan (a metacognitive strategy), and addressed the value of the task (a motivational aspect) were the most effective in enhancing student performance. Addressing the goal orientation of students (another motivational aspect), on the other hand, had a negative impact on the intervention effect. The intervention effect of strategy instruction was not related to student characteristics such as age, giftedness, special needs, or low socioeconomic status. To be able to control for the effects related to the instructional focus of the interventions, we took the strategies with a significant effect into account as covariates in the current meta-analysis. The influence of the subject domain and the measurement instrument were also examined in the meta-analysis by Donker et al. (2014). We summarized these findings in our study and analyzed them in a more in-depth manner.

The prior meta-research found large differences between the subject domains in which the interventions were implemented, although the effects were not always congruent. The results suggest that programs for reading comprehension have the lowest effect. In our study, we examined whether there were differences in intervention effect among the subject domains of reading, writing, mathematics, and science. Based on the findings of prior studies, we expected to find a lower effect for reading than for the other subject domains.

With respect to the implementer, we examined whether interventions implemented by a researcher indeed differed in effect from those implemented by a teacher. We also investigated the impact of computer-based interventions. We expected to find a higher effect for researcher-implemented programs than for teacher-implemented programs. However, we did not expect the average effect size of computer-based interventions to deviate much from the effect of the other programs.

Next, we analyzed the relationship between the duration and the intensity of the intervention and its effect on student performance. Although the research literature suggests that these factors are not important moderators of the effect size, this finding is not what one would expect. We would expect longer interventions to have a larger impact on student performance. We speculated that the counterintuitive findings in prior research might be the result of the researchers' preference for selecting a particular test to evaluate the intervention effect that matches the subject taught. Large interventions are probably tested using a test that covers a larger study domain than short interventions. We therefore hypothesized that duration and intensity would only affect the estimated effect size of an intervention if a test not specifically developed for the intervention was used as measurement instrument. This type of test is generally less task-specific and measures student performance in a broader sense than tests that are specifically designed to estimate particular intervention effects.

With regard to cooperative learning, prior research suggests that interventions in which students cooperate are more effective than those where there is no cooperation among students. We tested whether this finding could be reproduced in our study. We expected that aspects of the research method such as the implementation fidelity check and situation of the control group might have a small influence on the estimated intervention effect. Checking the implementation fidelity may lead to a higher effect, and the control group being aware of taking part in an experiment could result in a lower effect, as in these studies, novelty effects were taken into account. On the other hand, we did not expect to find a large effect of whether or not random assignment was applied, since we only included studies in our meta-analysis that took pretest scores into account.

Finally, the results of the prior research indicate that the measurement instrument used to evaluate the intervention's effectiveness influences the estimated effect. We expected that the effect would be higher when measured using a test specifically developed for the intervention than using an intervention-independent test. Self-developed tests are considered to focus on the (very) near transfer of the task learned and independent tests on the further transfer. We also investigated possible differential effects between standardized and unstandardized tests. Rosenshine (1994) has suggested that standardized tests are less sensitive to the measurement of intervention effects than other tests.

Method

Below, we briefly describe our search criteria for the initial study retrieval and indicate the eligibility criteria on which we based our choices regarding the inclusion of studies in our meta-analysis. In addition, we explain the study coding procedure and how we analyzed the data. For more information about the literature search we refer to the study by Donker et al. (2014) in which the same methods were used.

Literature Search

We started by searching the Internet databases, ERIC and PsycINFO. We decided to use a limited time span, from 2000 to 2011, concentrating on the most recent research studies. We took the year 2000 as starting point, because in that year, the *Handbook of Self-Regulation* by Boekaerts, Pintrich, and Zeidner was published, representing the beginning of a new era of research on this topic. The search terms we entered were "metacognit*" and "self-reg*," which had to form part of the titles of the articles. With respect to advanced search options, we limited our search to articles written in English and published in peer-reviewed journals (so books and book chapters were excluded from our analysis).

Eligibility Criteria

Our principles for inclusion were all based on our main criterion: empirical studies dealing with learning strategy instruction aimed at improving academic achievement. This criterion meant that we exclusively selected articles that included the dependent variable, academic achievement (operationalized as performance on one or more school subject domains). We excluded correlation studies that only examined the relationship between strategy use and student achievement. In these studies, learning strategy instruction is not implemented as training, so it is not possible to analyze the causal relationship between strategy instruction and student achievement. The other eligibility criteria included the following:

- With respect to the subject domain, the focus had to be on core academic subjects, such as reading, writing, mathematics, and science. Subjects such as music, arts, and physical education were excluded, as we were particularly interested in the effect of interventions on academic achievement in the core school domains.

- The research sample had to consist of primary or secondary school students up to and including the 12th grade, as in most European and the American school systems. This criterion was included to enable generalization of the results to school learning.

We also used the following methodological criteria to select studies of sufficient quality:

- The research design had to include a control group. If a control group were not included, it would be unclear if the results of the experimental group were explained by the intervention or by other factors.
- The research had to provide pretest and posttest measures. Studies that did not provide pretest scores were only included if it was indicated that there were no initial differences between the control group and the experimental group. We did not include studies with random assignment of students or classes to the experimental and the control group and which reported posttest data only, if they did not report that there were no initial differences between the experimental and the control groups: Use of random assignment does not necessarily mean that there were no pretest differences between the two groups.
- The study samples had to include at least 10 students per group. We used this criterion to ensure that the effect size (Cohen's *d*) would be approximately normally distributed (Hedges & Olkin, 1985).

Coding

The first two authors coded all articles. The intercoder reliability was based on 10 articles that these authors coded independently. With a 96% agreement, the intercoder reliability of the attributes was high. Table 1 presents an overview of the studies and their characteristics. We have highlighted the variables relevant to the current meta-analysis:

- *Subject domain*: We coded the subject domains in which the interventions were implemented, distinguishing five categories, including (comprehensive) reading, writing, mathematics, science, and other.
- *Implementer of the training*: Here we coded three categories, researcher or research assistant, teacher, and computer.
- *Duration of the intervention*: This was coded in weeks.
- *Intensity of the intervention*: This was coded as the number of sessions per week.
- *Intensity of each session*: This was coded as the duration of each session in minutes.
- *Cooperation*: This was decided in terms of whether the training focused on cooperative or on individual learning. We distinguished three categories: cooperation in intervention group and in control group, cooperation in intervention group but not in control group, and no cooperation in either group.

TABLE 1

Overview of the intervention characteristics

Authors	Subject	Implementer	*n* Weeks	*n* Sessions/ week	Minute/ session	Cooperation	Random assignment	Fidelity	Control group	Measurement instrument	*n* Experimental group	Hedges's *g*
Aleven and Koedinger (2002) (1)	M	PC	20	?	?	No	No	No, but	Exp	S	13	0.52
Aleven and Koedinger (2002) (2)	M	PC	20	?	?	No	No	No, but	Exp	S	22	0.64
Allen and Hancock (2008) (1)	R	T	10	2	10	No	Yes	No	BAU	SIT	45	0.14
Allen and Hancock (2008) (2)	R	T	10	2	10	No	Yes	No	BAU	SIT	60	0.24
Blank (2000)	O	T	13	?	?	No	Yes	No	Exp	S	24	−0.48
Boulware-Gooden, Carreker, Thornhill, and Malatesha (2007)	R	T	5	5	30	No	No	No	BAU	SIT and UIT	59	0.50
Bruce and Robinson (2001)	R	T	24	3	30	Ig	No	Yes	Exp	SIT	25	0.37
Brunstein and Glaser (2011)	W	(A)R	5	1	45	No	Yes	Yes	Exp	UIT	58	0.86
Camahalan (2006)	M	(A)R	6	5	60	Ig	Yes	No, but	BAU	UIT and S	30	0.54
Cantrell, Almasi, Carter, Rintamaa, and Madden (2010) (1)	R	T	40	5	55	No	Yes	Yes	BAU	SIT	171	0.25
Cantrell et al. (2010) (2)	R	T	40	5	55	No	Yes	Yes	BAU	SIT	194	0.09
De Acedo Lizarraga, de Acedo Baquedano, and Oliver (2010)	O	T	40	?	?	No	Yes	No	BAU	S	24	1.22
Dejonckheere, Van de Keere, and Tallir (2011)	S	(A)R	?	?	50	No	No	No	Exp	S	15	0.70
Dresel and Haugwitz (2008) (1)	M	PC	21	0.3	45	No	Yes	No, but	Exp	S	42	0.23
Dresel and Haugwitz (2008) (2)	M	PC	21	0.3	45	No	Yes	No, but	Exp	S	61	0.77
Erktin (2004)	M	T	?	?	?	No	No	No	BAU	S	21	0.87
García-Sánchez and Fidalgo-Redondo (2006) (1)	W	(A)R	8	3	50	Ig	Yes	Yes	BAU	S	48	2.54
García-Sánchez and Fidalgo-Redondo (2006) (2)	W	(A)R	8	3	50	Ig	Yes	Yes	BAU	S	41	1.96
Glaser and Brunstein (2007) (1)	W	(A)R	4	1	90	No	Yes	Yes	BAU	UIT and S	41	1.70
Glaser and Brunstein (2007) (2)	W	(A)R	4	1	90	No	Yes	Yes	BAU	UIT and S	34	0.93

(continued)

TABLE 1 (continued)

Authors	Subject	Implementer	*n* Weeks	*n* Sessions/ week	Minute/ session	Cooperation	Random assignment	Fidelity	Control group	Measurement instrument	*n* Experimental group	Hedges's *g*
Graham, Harris, and Mason (2005) (1)	W	(A)R	?	3	20	Ig	Yes	Yes	BAU	S	12	0.92
Graham et al. (2005) (2)	W	(A)R	?	3	20	Ig	Yes	Yes	BAU	S	12	1.03
Guterman (2003)	R	?	?	?	?	No	No	No	BAU	S	113	0.83
Harris, Graham, and Mason (2006) (1)	W	(A)R	26	3	20	Ig	Yes	Yes	BAU	S	21	0.71
Harris et al. (2006) (2)	W	(A)R	26	3	20	Ig	Yes	Yes	BAU	S	22	1.21
Hauptman and Cohen (2011) (1)	M	PC	21	1	15	No	Yes	No, but	Exp	UIT	41	0.35
Hauptman and Cohen (2011) (2)	M	(A)R	21	1	15	No	Yes	No, but	Exp	UIT	41	0.78
Huff and Nietfeld (2009) (1)	R	(A)R	2.4	5	35	No	Yes	No, but	BAU	SIT	21	0.00
Huff and Nietfeld (2009) (2)	R	(A)R	2.4	5	35	No	Yes	No, but	BAU	SIT	24	0.00
Jacobse and Harskamp (2009)	M	PC	?	2	30	No	No	No, but	BAU	S	23	0.76
Kaniel, Licht, and Peled (2000)	R	(A)R	5	3	35	No	Yes	No, but	Exp	UIT	70	0.66
Kapa (2007) (1)	M	PC	?	?	60	No	Yes	Yes	Exp	S	54	1.01
Kapa (2007) (2)	M	PC	?	?	60	No	Yes	Yes	Exp	S	57	0.97
Kapa (2007) (3)	M	PC	?	?	60	No	Yes	Yes	Exp	S	44	0.74
Kapa (2001) (1)	M	PC	9	1	60	No	Yes	Yes	Exp	S	110	0.60
Kapa (2001) (2)	M	PC	9	1	60	No	Yes	Yes	Exp	S	110	0.44
Kapa (2001) (3)	M	PC	9	1	60	No	Yes	Yes	Exp	S	110	0.30
Kim and Pedersen (2011)	S	PC	2	5	50	No	No	No, but	Exp	S	100	0.41
Kramarski and Dudai (2009) (1)	M	T	?	?	45	Ig and cg	No	Yes	BAU	UIT and S	32	0.24
Kramarski and Dudai (2009) (2)	M	T	?	?	45	Ig and cg	No	Yes	BAU	UIT and S	32	0.87
Kramarski and Gutman (2006)	M	T	5	4	90	Ig and cg	Yes	No	Exp	S	35	1.17
Kramarski and Hirsch (2003) (1)	M	T	20	1	60	No	Yes	No	Exp	?	20	0.80
Kramarski and Hirsch (2003) (2)	M	T	20	1	60	No	Yes	No	BAU	?	20	0.43

(continued)

TABLE 1 (continued)

Authors	Subject	Implementer	*n* Weeks	*n* Sessions/ week	Minute/ session	Cooperation	Random assignment	Fidelity	Control group	Measurement instrument	*n* Experimental group	Hedges's *g*
Kramarski and Mevarech (2003) (1)	M	T	2	5	?	Ig	Yes	Yes	BAU	S	105	0.55
Kramarski and Mevarech (2)	M	T	2	5	?	No	Yes	Yes	BAU	S	95	0.34
Kramarski and Mizrachi (2006) (1)	M	T	4	5	45	Ig and cg	Yes	No	Exp	S	20	1.80
Kramarski and Mizrachi (2006) (2)	M	T	4	5	45	Ig and cg	Yes	No	Exp	S	22	1.06
Kramarski and Ritkof (2002)	M	T	7	1	60	No	No	No	Exp	S	25	0.71
Kramarski and Zoldan (2008) (1)	M	T	13	3	45	No	Yes	Yes	BAU	S	32	0.67
Kramarski and Zoldan (2008) (2)	M	T	13	3	45	No	Yes	Yes	BAU	S	26	0.38
Kramarski and Zoldan (2008) (3)	M	T	13	3	45	No	Yes	Yes	BAU	S	30	0.76
Kramarski, Mevarech, and Arami (2002) (1)	M	T	6	5	45	Ig and cg	No	No	Exp	S	14	1.02
Kramarski, Mevarech, and Arami (2002) (2)	M	T	6	5	45	Ig and cg	No	No	Exp	S	17	1.17
Kramarski, Mevarech, and Lieberman (2001) (1)	M	T	?	?	?	Ig and cg	Yes	Yes	BAU	S	60	1.49
Kramarski, Mevarech, and Lieberman (2001) (2)	M	T	?	?	?	Ig and cg	Yes	Yes	BAU	S	60	0.56
Lubliner and Smetana (2005)	R	T	12	2	45	Ig	No	Yes	BAU	S	77	0.60
Mason (2004)	R	(A)R	?	?	20	Ig and cg	Yes	Yes	Exp	S	16	0.94
Mevarech and Kramarski (2004) (1)	M	T	4	?	40	Ig and cg	Yes	Yes	Exp	S	35	0.15
Mevarech and Kramarski (2004) (2)	M	T	4	?	40	Ig and cg	Yes	Yes	Exp	S	35	0.62
Meyer, Abrami, Wade, Aslan, and Deault (2010)	O	T	40	?	?	No	No	Yes	BAU	SIT	121	0.08
Michalsky, Mevarech, and Haibi (2009) (1)	S	T	17	1	?	Ig and cg	Yes	Yes	Exp	UIT and S	27	0.79
Michalsky et al. (2009) (2)	S	T	17	1	?	Ig and cg	Yes	Yes	Exp	UIT and S	27	0.62
Michalsky et al. (2009) (3)	S	T	17	1	?	Ig and cg	Yes	Yes	Exp	UIT and S	27	1.62

(continued)

TABLE 1 (continued)

Authors	Subject	Implementer	*n* Weeks	*n* Sessions/ week	Minute/ session	Cooperation	Random assignment	Fidelity	Control group	Measurement instrument	*n* Experimental group	Hedges's *g*
Molenaar, Chiu, and Sleegers (2011) (1)	W	PC	?	?	60	Ig and cg	Yes	No, but	Exp	S	18	0.45
Molenaar et al. (2011) (2)	W	PC	?	?	60	Ig and cg	Yes	No, but	Exp	S	18	0.85
Mourad (2009)	W	T	?	3	45	Ig	Yes	Yes	BAU	S	34	2.54
Pennequin, Sorel, Nanty, and Fontaine (2010)	M	(A)R	5	1	60	No	Yes	No	Exp	UIT	24	2.17
Perels, Dignath, and Schmitz (2009)	M	T	3	3	?	Ig	No	No	Exp	S	26	0.45
Perels, Gurtler, and Schmitz (2005) (1)	M	T	6	1	90	No	Yes	No	BAU	S	59	0.23
Perels, Gurtler, and Schmitz (2005) (2)	M	T	6	1	90	No	Yes	No	BAU	S	59	0.33
Perels, Gurtler, and Schmitz (2005) (3)	M	T	6	1	90	No	Yes	No	BAU	S	59	0.46
Peters and Kitsantas (2010a)	S	T	?	?	30	No	Yes	Yes	BAU	UIT and S	69	0.69
Peters and Kitsantas (2010b)	S	T	?	?	?	No	Yes	Yes	BAU	UIT and S	41	0.74
Reynolds and Perin (2009) (1)	W	T	5	1	45	No	Yes	Yes	Exp	S	40	1.25
Reynolds and Perin (2009) (2)	W	T	5	1	45	No	Yes	Yes	Exp	S	39	0.92
Souvignier and Mokhlesgerami (1)	R	T	20	1	45	No	No	Yes	BAU	SIT	95	0.14
Souvignier and Mokhlesgerami (2)	R	T	20	1	45	No	No	Yes	BAU	SIT	146	−0.10
Souvignier and Mokhlesgerami (3)	R	T	20	1	45	No	No	Yes	BAU	SIT	89	0.12
Souvignier and Mokhlesgerami (4)	R	T	20	1	45	No	No	Yes	BAU	SIT	115	0.49
Souvignier and Mokhlesgerami (5)	R	T	20	1	45	No	No	Yes	BAU	SIT	64	0.57
Stoeger and Ziegler (2008)	M	T	5	?	?	No	Yes	No	BAU	S	112	0.36
Stoeger and Ziegler (2010)	M	T	5	?	?	No	Yes	No	BAU	S	100	0.48
Tajika, Nakatsu, Nozaki, Neumann, and Maruno (2007) (1)	M	T	1	1	?	No	No	No	BAU	S	27	1.04
Tajika et al. (2007) (2)	M	T	1	1	?	No	No	No	BAU	S	26	0.04

(continued)

TABLE 1 (continued)

Authors	Subject	Implementer	*n* Weeks	*n* Sessions/ week	Minute/ session	Cooperation	Random assignment	Fidelity	Control group	Measurement instrument	*n* Experimental group	Hedges's *g*
Talebi (2009) (1)	R	(A)R	?	?	60	No	No	No, but	BAU	UIT and S	30	1.14
Talebi (2009) (2)	R	(A)R	?	?	60	No	No	No, but	BAU	UIT and S	30	1.46
Teong (2003)	M	PC	2	2	60	Ig and cg	No	No, but	Exp	S	20	0.59
Torrance, Fidalgo, and García (2007)	W	T	10	1	66	No	No	Yes	BAU	S	71	2.54
Tracy, Reid, and Graham (2009)	W	T	?	?	?	Ig	Yes	Yes	BAU	S	63	0.34
Van Keer and Vanderlinde (2010) (1)	R	(A)R	40	1	50	Ig	No	No	BAU	SIT	308	0.05
Van Keer and Vanderlinde (2010) (2)	R	(A)R	40	1	50	Ig	No	No	BAU	SIT	277	0.03
Vaughn, Klingner, and Swanson (2011)	R	T	18	2	50	Ig	Yes	Yes	BAU	SIT	400	−0.01
Wright and Jacobs (2003)	R	T	20	2	30	Ig	No	No	BAU	SIT	25	0.68
Zion, Michalsky, and Mevarech (2005) (1)	S	T	12	?	?	Ig and cg	Yes	Yes	Exp	S	102	0.65
Zion et al. (2005) (2)	S	T	12	?	?	Ig and cg	Yes	Yes	Exp	S	100	0.76

Note. Several studies tested the effects of multiple interventions. Each intervention is listed on a separate row in the table. For subject: R = Reading; W = Writing; M = Math; S = Science; O = other subject. For implementer: A/R = assistant/researcher; T = teacher. For cooperation: Ig = intervention group; cg = control group. For situation of control group: BAU = business as usual; Exp = idea of experiment (not business as usual). For measurement instrument: S = self-developed; UIT = unstandardized intervention-independent test; SIT = standardized intervention-independent test.

- *Random assignment*: This was coded as "1" when students or classes were randomly assigned to the experimental group and the control group and coded as "0" otherwise.
- *Fidelity of the implementation*: We distinguished three categories: studies that reported that the fidelity of the implementation was checked and was good (fidelity yes), studies in which there was no mention of a fidelity check (fidelity no), and (3) studies in which the fidelity was not checked but in which the researcher himself or the computer was the implementer (fidelity no but). We distinguished this last category as we assumed that when the researcher or the computer implemented the intervention, there was less chance of the intervention not being implemented as it should be compared with if the intervention was implemented by the teachers or the research assistants.
- *Situation of the control group*: This was coded as "0" if the control group had their normal lessons (business as usual) and coded as "1" if the teacher or the students perceived that the lessons were not as usual and were part of an experiment. It was not always clear what the situation of the control group was. To reduce interpretation differences, only one of the authors coded for this variable.
- *Measurement instrument*: We distinguished three types of instruments used to assess the effects of the interventions: tests developed by the researcher himself to test the effect of the intervention (further referred to as self-developed tests), unstandardized tests developed independently of the intervention, and standardized tests (which were also developed independently of the intervention). We did not distinguish between treatment-inherent and treatment-independent measures, as Slavin and Madden (2011) did, because all studies in our meta-analysis tested an instructional change. In experiments where instructional programs are tested, as in our case, treatment-inherent performance tests are not commonly used.

Analysis

Meta-analysis is a statistical technique by which the quantitative results of multiple studies focused on one particular research question are combined. In a metastudy, the unit of analysis is not the individual participant, as is the case in primary studies, but the effect size determined on the basis of these primary studies' outcomes. A meta-analysis enables one to systematically review multiple studies on the same subject. The summary effect can be calculated based on all studies included in the meta-analysis. Furthermore, it can be examined if there are moderators that influence the size of the effect. In a meta-analysis, each primary study is assigned a different weight, depending on the precision with which the effect size has been measured. In computing a summary effect, effect sizes measured with greater precision are, therefore, given more weight. In general, studies with larger sample sizes are measured with greater precision. The exact weight assigned to each study is the inverse of the variance (1/variance).

To perform our meta-analysis, we used the statistical packages, Comprehensive Meta-Analysis (CMA) Version 2, developed by Biostat (see www.meta-analysis.com), and Hierarchical Linear Modeling (HLM) Version 6, developed by Raudenbush, Bryk, and Congdon (2004). Because the interventions included in the meta-analysis differed in many respects, we used a random effects model to estimate the weighted average effect size. To prevent extreme effect sizes of individual studies from influencing the results in an unrepresentative way, we adjusted these values through Windsorizing (Lipsey & Wilson, 2001). Outliers were recoded to the general unweighted mean of the effect sizes plus or minus two times the standard deviation. Because of this, the results should more strongly reflect the average tendencies of the interventions included in the meta-analysis.

CMA was used to compute the effect size and the variances of the individual interventions. First, the standardized mean difference *d* was computed based on the pretest and posttest scores of the experimental and the control group. Because *d* tends to overestimate the effect in small sample sizes, we had the program convert *d* into Hedges's *g*, which is the unbiased estimate of the population standardized mean difference. CMA was also used to calculate average weighted effect sizes and analyze publication bias. We applied hierarchical linear modeling to perform a metaregression analysis with multiple predictors. A metaregression is like a normal regression analysis, except that in the former, the predictors or moderators are categorized at the level of the interventions and the dependent variable is the size of the effect of these interventions.

We used metaregression analysis to examine the influence of the attributes on the intervention effect size. For each attribute, we tested a model using the intervention effect size as criterion, the attribute as predictor, and the significant learning strategies as covariates. We included these strategies to correct for the effect related to the instructional focus of the interventions. The significant strategies were general metacognitive knowledge, planning, task value, and goal orientation, as indicated in the study by Donker et al. (2014). The resulting regression equations were of the following type:

$$\begin{aligned}\text{Effect size} = \text{intercept} &+ B_{\text{general metacognitive knowledge}} + B_{\text{planning}} \\ &+ B_{\text{task value}} + B_{\text{goal orientation}} + B_{\text{attribute}}\end{aligned}$$

In the results section, we only report the effects of the covariates in the final analyses in which we examined the joint effects of multiple attributes.

Results

Description of the Studies Included in the Meta-Analysis

The literature search resulted in 58 studies that met our eligibility criteria. Several of these studies reported the effects of multiple intervention types, with a unique experimental group for each intervention. The resulting number of interventions (and consequently also the number of different experimental groups) in our meta-analysis was 95. About half of the interventions focused on primary school students and the other half on secondary school students (the average grade level was 6.4, $SD = 2.2$). The interventions were conducted in various countries: the majority were in the Middle East (38.9%; mostly in Israel), followed by Europe (30.5%), North America (24.2%), Asia (5.3%), and Australia (1.1%). The studies were published between 2000 and 2011. The average number of students in the experimental group was 60.1 ($SD = 61.8$; minimum = 12; maximum = 400).

Of these 95 interventions, 23 focused on reading comprehension, 16 on writing, 44 on mathematics, 9 on science, and 3 on other subjects. Almost two thirds of the interventions were implemented by the teacher, about 22% by the researcher or the researcher's assistant, and another 15% via the computer. On average, instruction took 13 weeks, although the variability in duration was quite large (*SD* = 10.7 weeks; minimum = 1 week; maximum = 40 weeks). The average number of sessions per week was 2.3 (*SD* = 1.6; minimum = 0.3; maximum = 5), and the average duration of each session was 48 minutes (*SD* = 18.7 minutes; minimum = 10 minutes; maximum = 90 minutes). Unfortunately, 22 studies did not report the interventions' duration, 29 did not indicate their intensity, and 23 did not provide information on the duration of a session. In 37 programs, the students in the intervention group cooperated with one another. In 20 of these programs, the control group was also allowed to cooperate.

The effectiveness of the 95 interventions on student performance was measured by administering a total of 180 tests. The majority of these tests (122) were self-developed; 50 tests were developed independently of the intervention, of which 30 were standardized tests. No information was provided for 8 tests. In almost all cases, the research design was a pretest-posttest control group approach. Only 5 interventions used a posttest-only control group design (but indicated that there were no initial differences between the experimental group and the control group) and 4 interventions used both designs. Sixty-four of the 95 interventions randomly assigned students to the experimental and control groups, and implementation fidelity was checked in 50 interventions. In 17 cases, there was no fidelity check, but the intervention was implemented by the researcher himself, or by the computer. Finally, for 55 interventions, we coded the situation of the control group as business as usual; in the other interventions, the students in the control group perceived that they were taking part in an experiment.

We computed the summary effect of all 95 interventions, resulting in an average (weighted) effect size (Hedges's g) estimate of 0.66 (SE = .05, confidence interval [CI] = 0.56–0.76). Following Cohen (1988), this is a medium to high effect size. The Q statistic indicates that there was significant heterogeneity among the effect sizes (Q = 439.3; df = 94; p = .000), which means that it is unlikely that all interventions shared the same true effect size. The I^2 was 78.6, which suggests that 78.6% of the dispersion of the interventions' effect sizes reflects real differences in true effect size and that only 21.4% was due to random error. The estimated variance of the true effect sizes (T^2) was .17.

We also tested whether there was any publication bias. We used Duval and Tweedie's trim and fill method for a random effects model to see if any studies were missing in the meta-analysis (Borenstein, Hedges, Higgins, & Rothstein, 2009; Peters, Sutton, Jones, Abrams, & Rushton, 2007), but this seemed not to be the case. It is unlikely, therefore, that the results were significantly affected by publication bias. This result was supported by the classic fail-safe *N*s of Rosenthal (1979), which was 5,196, and Orwin (1983), which was 777, indicating that 5,196 and 777 interventions, respectively, would have been needed to be added to the meta-analysis, with an effect size of 0, before the effect found would have become nonsignificant at $p < .05$.

TABLE 2

Mean effect size per subject domain and metaregression

Subject	*n* Interventions	Mean Hedges's *g* (*SE*)	Regression, *B* (*SE*)
Intercept			.17 (.08)*
Reading	23	0.36 (.08)**	
Writing	16	1.25 (.12)**	.65 (.13)**
Math	44	0.66 (.06)**	.32 (.08)**
Science	9	0.73 (.13)**	.33 (.12)**
Other	3	0.23 (.23)	.06 (.20)

Note. In the regression analysis, the reference category for the subject domain was "reading." The significant strategies were included as covariates in the regression model; the mean Hedges's *g* was computed without these covariates.
**$p < .01$. *$p < .05$.

Effects of the Attributes

Subject Domain

Most interventions in our meta-analysis were directed at one of the four subject domains: reading comprehension, writing a text, mathematics, and science. Table 2 depicts the average effect estimated for each subject domain (see the column with Hedges's *g;* in this analysis, we did not correct for the learning strategies). Table 2 shows that learning strategy instruction has the largest average effect on student performance for the subject writing a text. With a Hedges's *g* of 1.25, this effect is very high. For the other subjects, the effect size is much lower, although it is still considerable for science and mathematics. For reading comprehension, the effect of strategy instruction on student performance is, on average, small to moderate. The three interventions applying to other subjects have no significant effect on student performance. An analysis of variance for meta-analytical data proved that the between-subject differences were significant, Q-between(4) = 27.6, $p < .01$.

To examine the between-subjects differences more thoroughly, we performed a metaregression analysis. This allowed us to compare two separate groups with each other rather than analyzing the between-groups differences as a combined whole. In addition, we were able to correct for the effects that related to the learning strategies taught by including them as covariates. The findings of the analysis revealed that the average effect of strategy instruction was significantly higher in writing than in all other subjects. Furthermore, the average effect of the instruction in reading comprehension was significantly lower than in writing, mathematics, and science but not so compared with the other-subject interventions. The effects for mathematics, science, and the category other subjects did not differ significantly from each other. The last column of Table 2 shows the metaregression results with the subject domain reading comprehension as the reference category. In comparison with the regression model with covariates only, the subject domain explained 33.1% of the variance in effect size.

TABLE 3

Mean effect size per implementer and metaregression

Implementer	*n* Interventions	Mean Hedges's *g* (*SE*)	Regression, *B* (*SE*)
Intercept			.49 (.11)**
(Assistant) researcher	21	0.93 (.15)**	
Teacher	57	0.60 (.06)**	−.20 (.10)*
PC	16	0.55 (.06)**	−.02 (.14)

Note. In the regression analysis, the reference category for the implementer was "assistant/researcher." The significant strategies were included as covariates in the regression model; the mean Hedges's *g* was computed without these covariates.
$^{**}p < .01$. $^{*}p < .05$.

Implementer of the Intervention

The meta-analysis included 21 interventions implemented by the researchers or their assistants. In 17 of these interventions, the control group was taught by the regular teacher. Of the 16 interventions in which the computer provided the instruction, in 15 the control group also received instruction from the computer. The implementer of one intervention was unknown. Table 3 provides the results of the analysis of the implementer's influence on the effectiveness of the intervention. It appeared that interventions implemented by the researcher or the researcher's assistant had the largest effect. The average effect of interventions implemented by the teacher was about the same as those provided via the computer. A meta-analysis of variance, however, indicated that there were no significant between-groups differences, Q-between(2) = 5.2, $p = .07$. For the interventions implemented by the researcher or assistant, we also examined whether there were differences related to the instructor of the control group. It appeared, however, that the effect size was not influenced by whether the control group was also taught by the assistant/researcher or the regular teacher, Q-between(1) = 0.6, $p = .43$.

We examined the group differences more thoroughly using metaregression. In this analysis, the learning strategies served as covariates. In this way, we could correct for differences related to the instructional focus of the interventions. The analysis revealed significant differences in effect between interventions implemented by the assistant or the researcher and those implemented by the teacher or using a computer. There were no differences in effect between interventions implemented by the teacher and those provided via the computer. The implementer explained 6.6% of the variance in effect size.

Duration of the Intervention

We found a small effect of the duration of the intervention on its effectiveness. Longer interventions had a slightly smaller effect on student performance than shorter ones. Metaregression analysis with the learning strategies as covariates and duration as predictor reported an unstandardized regression coefficient of $B = -.01$ ($SE = .00$) for the number of weeks that the intervention took. This finding implies that an intervention with a duration of 10 weeks had on average a 0.1

TABLE 4

Mean effect size for cooperation and metaregression

Cooperation	*n* Interventions	Mean Hedges's *g* (*SE*)	Regression, *B* (*SE*)
Intercept			.41 (.07)**
Cooperation in ig but not in cg	17	0.73 (.14)**	−.23 (.10)*
Cooperation in ig and cg	20	0.84 (.09)**	
No cooperation in ig or cg	58	0.58 (.06)**	

Note. In the regression analysis, the reference category for cooperation was "intervention group and control group both do or do not cooperate." The significant strategies were included as covariates in the regression model; the mean Hedges's *g* was computed without these covariates. Ig = intervention group; cg = control group.
$^{**}p < .01$. $^{*}p < .05$.

higher effect size than an intervention of 20 weeks. The duration of the intervention explained 20.2% of the variance in effect size.

Intensity of the Intervention

Metaregression analysis with the learning strategies as covariates and intensity as predictor showed that the intervention's intensity, measured as the number of sessions per week, had no significant influence on its effectiveness. The regression coefficient for intensity was $B = -.01$ ($SE = .01$). The intensity of the intervention explained no variance in effect size.

Intensity of Each Session

The duration of the sessions had a small influence on the intervention effect. The metaregression showed a very small but significant regression coefficient for this attribute ($B = .0045$; $SE = .00$). Interventions with more intensive sessions had a slightly higher effect than interventions with less intensive sessions. The regression coefficient indicates that a difference in session duration of 30 minutes resulted in an effect size difference of 0.14. The duration of the sessions explained no variance in effects, however.

Cooperation During the Intervention

Table 4 presents the average effect size for interventions in which both the intervention group and the control group cooperated, those in which neither group cooperated, and those in which only the intervention group cooperated. The table shows that there are some small differences among the groups which are significant, Q-between(2) = 6.0, $p = .05$. We were especially interested in whether cooperation had a positive influence on the effectiveness of the interventions. We therefore compared the effects of the programs in which the intervention group cooperated but the control group did not with the effects of the other programs. We used a metaregression analysis with the learning strategies as covariates. The last column of Table 4 displays the results. It appeared that interventions in which students were allowed to cooperate had a lower effect

TABLE 5

Mean effect size for implementation fidelity and metaregression

Implementation fidelity checked	*n* Interventions	Mean Hedges's *g* (*SE*)	Regression, *B* (*SE*)
Intercept			.36 (.08)**
Fidelity yes	50	0.71 (.07)**	
Fidelity no but	17	0.58 (.09)**	.06 (.11)
Fidelity no	28	0.61 (.09)**	.05 (.09)

Note. In the regression analysis, the reference category for implementation fidelity was "fidelity yes." The significant strategies were included as covariates in the regression model; the mean Hedges's *g* was computed without these covariates.
**$p < .01$. *$p < .05$.

on average than those without cooperation. Cooperation explained 11.2% of the variance in intervention effect.

Random Assignment

The average effect for intervention studies with random assignment was higher ($g = 0.70$, $SE = .06$) than for interventions without random assignment ($g = 0.58$, $SE = .09$). This difference was not significant though. The findings of the metaregression analysis, in which we corrected for the significant learning strategies, also showed that there was no difference in effect between interventions in which students or classes were randomly assigned to the experimental and control groups and those without random assignment. The regression coefficient for random assignment was .02 ($SE = .08$; the reference category was "no random assignment"). This attribute did not explain any variance in the intervention effect.

Fidelity of the Implementation

Table 5 shows the results of the moderator analysis for the implementation fidelity. It appeared that there were no differences in effect related to the fidelity of the intervention implementation. This attribute did not explain any variance.

Situation of the Control Group

Interventions in which the control group's situation was business as usual had an average effect of 0.61 ($SE = .07$). Interventions with a different situation for the control group had an average effect of 0.72 ($SE = .06$). This difference was not significant. However, the metaregression analysis with the learning strategies as covariates did show a difference in effect. The effect for the studies with the business as usual control group was 0.23 ($SE = .08$) lower than for the other studies. The situation of the control group variable explained 14.8% of the variance in the intervention effect.

Measurement Instrument

The findings of a metaregression analysis with the learning strategies as covariates and the measurement instrument as the predictor showed that self-developed tests yielded a 0.25 (*SE* = .08) higher effect size than intervention-independent tests; the difference is significant. The measurement instrument explained 19% of the variance in the effect size. Chiu's (1998) research indicated that standardized tests yield a lower effect size than unstandardized tests. We therefore also checked for these differences. We found that standardized tests yielded an effect size that was 0.43 (*SE* = .07) lower than that for unstandardized tests. Distinguishing between these two types of measurement instrument explained 52.1% of the variance in the intervention effect.

The Simultaneous Analysis of the Attributes

Thus far, we examined the effect of each individual attribute on student performance. Next, we addressed the effects of the significant attributes simultaneously. Table 6 reports the results of our metaregression analysis with the learning strategies as covariates. Only the difference between reading and writing interventions and writing and math interventions remains significant, as does the effect of the intensity of the sessions. The difference between the assistant/researcher as implementer of the intervention and the computer as implementer became significant. The effects of the other attributes were not significant, presumably because of the large number of variables in the model in combination with a modest number of cases (interventions). The significant attributes together explained 55.1% of the variance in intervention effects compared with the model with only the learning strategies as covariates (the variance of the covariates-only model was .107. As an aside, the learning strategies explained 40.5% of the variance compared with the empty model [without variables]. The variance of the empty model was .181).

We then performed a forward regression analysis to examine if a model with fewer attributes would yield a different picture. We started with the model with the covariates only, and in every step, we included the attribute that explained most variance. This resulted in a model with four attributes. Adding more attributes did not improve the model fit in terms of percentage explained variance in intervention effect. Table 6 displays this forward regression model. This model explained 63.2% of the variance in the intervention effect compared with the model with only the learning strategies as covariates. It shows that interventions implemented by the assistant/researcher had a higher effect than those implemented by the teacher or the computer. The difference in effect between teachers and computers was not significant. Furthermore, the duration of a session still slightly influenced the intervention effect. A 30-minute difference in duration resulted in an average effect size difference of 0.12, with longer sessions being a little more effective than shorter ones. The effect of cooperation was not significant ($p = .052$) but suggests that allowing cooperation between students during an intervention reduces its effect. Finally, the difference in effect between standardized and unstandardized tests is still present.

Because forward regression analysis, like other stepwise regressions, has the disadvantage that it yields results that are not replicable (Lewis, 2007; Thompson, 1995), we examined whether two other models, one containing all implementation attributes and one containing all attributes related to the study method, would produce comparable results. Table 6 also displays these models and shows that

TABLE 6

Metaregression with multiple attributes

Variable	Significant attributes, *B* (*SE*)	Forward regression, *B* (*SE*)	Implementation attributes, *B* (*SE*)	Study method attributes, *B* (*SE*)
Intercept	.47 (.17)**	.53 (.12)**	.29 (.14)*	.64 (.11)**
General metacognitive knowledge	.19 (.08)*	.20 (.07)**	.21 (.08)*	.21 (.07)**
Planning	.21 (.07)**	.21 (.07)**	.20 (.08)*	.14 (.07)
Task value	.64 (.23)**	.81 (.20)**	.61 (.23)**	.88 (.20)**
Goal orientation	−.19 (.14)	−.21 (.11)	−.26 (.15)	−.26 (.12)*
Subject				
Writing	.32 (.16)*		.53 (.14)**	
Math	.07 (.14)		.28 (.10)**	
Science	.12 (.16)		.34 (.13)*	
Other	−.06 (.20)		−.07 (.35)	
Implementer				
Teacher	−.16 (.09)	−.17 (.08)*	−.20 (.09)*	
PC	−.28 (.14)*	−.28 (.12)*	−.27 (.14)*	
Duration of intervention	−.00 (.00)		−.00 (.00)	
Intensity of intervention			.01 (.01)	
Intensity of each session	.004 (0.002)*	.004 (.001)**	.003 (.002)*	
Coop in ig but not in cg	−.18 (.09)	−.16 (.08)	−.18 (.09)	
Random assignment				−.09 (.07)
Fidelity				
Fidelity no but				−.04 (.10)
Fidelity no				.02 (.08)
Cg not as usual	.05 (.09)			.07 (.08)
Standardized test	−.30 (.15)	−.45 (.08)**		−.42 (.09)**
Variance	.048	.039	.057	.054

Note. The reference category for the strategies is "strategy not included in intervention"; for the subject "reading"; for the implementer "assistant/researcher"; for cooperation "intervention group and control group both do or do not cooperate"; for random assignment "yes"; for check of implementation fidelity "yes"; for the situation of the control group "business as usual"; and for the measurement instrument "unstandardized test." Ig = intervention group; cg = control group.
$^{**}p < .01$. $^{*}p < .05$.

many of the same attributes appear to have a significant influence on the intervention effect as were included in the forward regression model. The model containing the implementation attributes shows significant effects for subject domain, the implementer, and the intensity of each session and an almost significant effect for cooperation (p = .051). The model containing the attributes related to the study method shows a significant effect for the measurement instrument only. The final forward regression model, however, explains more variance in intervention effect than do these two additional models.

Discussion

In this meta-analysis, we examined whether the effect of an educational intervention on students' academic performance was influenced by (a) the attributes of the intervention related to the implementation and (b) the attributes related to the method used to examine the effect of the intervention. Prior meta-analyses have suggested that these attributes do matter. However, in almost all of these studies, the investigation of these effects was merely a side issue. And more important, they did not always produce the same results. Through our meta-analysis, we aimed to offer a clearer picture of the effects of these attributes. Although this study included 95 learning strategy instruction interventions, we believe that its findings are also relevant to other educational instructions and their effect studies. The generalizability to interventions in educational settings with a nonacademic focus, such as students' behavioral problems, motivation, or classroom management, may be problematic, because the goals of these types of interventions are different from those of instructional interventions. However, we believe that the findings can be generalized to educational interventions with different instructional content, but focusing on similar goals, namely, enhancing students' academic performance. The attributes examined in the current meta-analysis are common characteristics of instructional interventions and are not unique to learning strategy instructions. Therefore, we assume that the findings also apply to other types of educational instruction interventions and argue that the results of this meta-analysis can be used to improve future interventions, enabling these to be implemented under the best conditions and with good experimental designs in which to examine their effects.

We examined the influence of the following attributes: the subject domain in which the intervention was implemented, the implementer, the duration of the intervention, the intensity of the intervention (number of sessions per week and duration of each session), whether cooperation was included, the random assignment of students to the experimental group and the control group, the implementation fidelity, the situation of the control group, and the type of measurement instrument used to evaluate the intervention. The findings of the analyses, in which we examined the effect of each attribute separately, indicated that except for the attributes random assignment, implementation fidelity, and number of sessions per week, all attributes were significantly related to the intervention effect. Analyzing the effects of the significant attributes simultaneously resulted in a model with too many predictors relative to the number of interventions included in the meta-analysis. We therefore executed a forward regression analysis to obtain a model with fewer predictors but with better interpretable results. We recognize that forward regression analysis has several disadvantages, as discussed by Lewis (2007) and Thompson (1995). Lewis (2007) suggested that a better method for model building is to determine the order in which the predictors are included in the regression analysis based on theory and prior findings. In our view, however, there was no logical order of predictors for the current study. To verify the forward regression model results, we tested two additional models, one for each group of attributes.

These models yielded many of the same significant predictors as were included in the final model of the forward regression analysis. In addition to the learning strategies as covariates, the final forward regression model included the attributes "standardized versus unstandardized measurement instrument," the implementer of the intervention, student cooperation, and the duration of each session. These variables were, therefore, likely to be the most important predictors of the intervention effect. Together, they explained 63.2% of the variance in intervention effect size, indicating that it is important to take into account not only the aspects related to the instructional focus of an intervention but also these attributes in order to optimize the implementation of the intervention and its effect study. We now discuss the effect of each attribute we examined.

We first examined the influence of the subject domain in which the instruction was provided. It appeared that interventions in reading comprehension, writing, math, and science significantly enhanced student performance but that there were large differences in their effect size. The simultaneous analysis of all significant attributes indicated that part of these differences related to other attributes, as the differences in effect between the subject domains were reduced when we took the other significant attributes into account. These results suggested that writing interventions had a greater effect than interventions in the other subject domains. The forward regression analysis, however, showed that the subject domain was not an important moderator of the intervention effect. On the basis of the findings of prior meta-analyses, we expected to find a lower effect for interventions in reading comprehension. We did find this result, but only in the analyses in which we did not take the attributes related to the study method into account. A possible explanation for the finding that the effect for writing is higher than that for the other subjects may be that this subject is taught less frequently and explicitly to students than the other subjects. Students' performance in writing might, therefore, be more susceptible to instruction than in the other subject domains.

With respect to the implementer of the intervention, our results revealed that instruction provided by the assistant or the researcher produced a greater effect than that provided by the teacher or the computer. The teacher and the computer were equally effective in implementing an intervention. Prior meta-analyses have also indicated that researcher-implemented interventions have a larger effect than instructions provided by the teacher (Chiu, 1998; Dignath & Büttner, 2008). We hypothesized that interventions implemented by the assistant or the researcher would have a higher effect than those implemented by the teacher because the latter would perhaps be less motivated and would not implement the intervention properly, would be less confident, and would be less inclined to "teach to the test" than researchers. Another possible reason is the novelty effect that occurs when the researcher is the implementer. In our analysis, we could only check whether the implementation fidelity influenced the intervention effect and whether novelty effects occurred; that is, whether students put more effort into their work because they found it interesting that a person other than the regular teacher came to teach them. Our findings indicated that studies in which the implementation fidelity was checked did not differ in their effect from studies in which this was not checked. Therefore, it is not likely that differences in implementation fidelity between researchers and teachers explain the influence of the implementer on the effect size. We did expect to find some effect of implementation fidelity, though. Apparently, however, checking implementation fidelity was not necessary for the intervention to be a success. The findings of prior research indicate that teachers do not always implement programs properly (Datnow & Castellano, 2000). Even if this was the case in the studies we included in the meta-analysis, teachers may still have implemented the critical components.

We examined the possibility of novelty effects by comparing the effect in two situations, one in which both the experimental and the control groups were taught by the assistant or the researcher and one where the experimental group was taught by the assistant or the researcher and the control group by the regular teacher. If novelty effects occurred, they would only be visible in the second situation. The results, however, showed that the average effect sizes in the two situations did not differ significantly. In addition, we examined whether the estimated intervention effect was lower when the control group perceived that they were taking part in an experiment instead of having normal lessons. In these studies, novelty effects could no longer explain the intervention effect. However, the results did not indicate the presence of a novelty effect. Therefore, novelty effects were not the explanation for the greater effect of interventions implemented by the assistant or the researcher. Actually, we initially found that the effect was lower in studies in which the control group had normal lessons instead of an experiment-like situation. However, in the simultaneous analysis of the significant attributes, this effect became almost zero and was no longer significant. We should note that the result with respect to the situation of the control group is not of high confidence. We agree with Chiu (1998) that coding this aspect was often not easy, as in many articles it was not clearly described whether the control group had normal lessons or whether there was a special, experiment-like, situation.

We also analyzed the effect of the interventions' duration (measured in weeks) and found a very small relationship between their duration and their effectiveness. Longer interventions had a slightly lower effect than shorter ones. Moreover, when the other attributes were taken into account, the effect of the duration of the intervention became nonsignificant. These results largely resemble the findings of the prior meta-analyses. However, the finding that an intervention's duration has little or no influence on its effectiveness is a remarkable one, to say the least. A possible explanation is that researchers tend to choose a measurement instrument that best fits the task: Since short interventions are likely to focus on a relatively specific task, the effect is likely to be estimated using a test suitable for measuring this particular item. In contrast, longer interventions are likely to be aimed at less specific tasks, while their effect is probably estimated using a test that measures student performance in a broader manner. The results of the simultaneous analysis of the significant attributes, however, indicated that this hypothesis is not true. In this analysis, we included the type of measurement instrument as predictor and found no effect of the intervention's duration. An additional analysis with the measurement instrument and the duration as the only predictors produced the same finding. It also did not seem as if short interventions had a narrower instructional focus than longer interventions. We compared the number of learning strategies addressed in the interventions but found no association with the interventions' duration.

The intensity of the intervention, the number of sessions per week, did not affect its effectiveness, but the duration of each session did to a very small extent. The sessions in this meta-analysis lasted between 10 and 90 minutes, and within this range, longer sessions were somewhat more effective than shorter ones. This finding can be explained by the fact that students need appropriate time to master new skills and that short training sessions may not be sufficient for students who need more learning time. Also, very short sessions may not allow teachers enough time to provide students with individualized feedback and correctives. These factors are important components of effective instruction (Creemers, 1994; Guskey, 2007).

Next, we examined whether cooperative learning had a positive influence on the effectiveness of the interventions, but this was not the case. The contrary was true: Cooperation negatively affected the intervention effect. Among other researchers, Slavin (1991) reviewed the large amount of research on cooperative learning, concluding that this concept is an effective teaching method to enhance student performance. Our finding that cooperation adds no value to the effectiveness of learning strategy interventions is, therefore, not what we would have expected. One possibility is that learning strategy instruction is so effective in itself that cooperative learning cannot add any more value to it. Another explanation can be found in Slavin's review: Cooperative learning can take several forms that differ in effectiveness. Perhaps cooperative learning was not implemented very successfully in the interventions included in our meta-analysis, as it was not the major focus of the researchers.

Slavin (1991) reported that to enhance student performance, cooperative learning should include the two elements, group goals and individual accountability. Groups have to work together to reach a goal, and the group's evaluation depends on the performances of all group members. Most interventions in the current meta-analysis did not contain these elements. The meta-analysis by Dignath and Büttner (2008), however, did show a positive effect of cooperative learning in learning strategy instruction. Although the authors found no positive effect of cooperative learning for interventions implemented in primary schools, they did observe positive impacts of cooperation in secondary school interventions. In an additional analysis, we checked if differences in effect could be found between primary and secondary education, but this was not the case. An explanation for the mixed results of the current meta-analysis and that of Dignath and Büttner (2008) might be that we also checked whether the students in the control group cooperated. It appeared that in more than half of the interventions where the students in the intervention group cooperated, the control group also did. This situation might have also been the case in the other meta-analysis, some of whose results were possibly influenced by this condition (although then, no significant effect for cooperation would have been more likely).

We then examined the effect of the application of random assignment, but as we expected, this attribute was not related to the intervention effect. This finding is likely to be because all studies included in the meta-analysis corrected for pretest differences between the experimental and the control groups. Finally, we investigated the influence of the type of measurement instrument used to evaluate the interventions' effectiveness. We examined the differences between tests developed by the researcher himself, especially designed to estimate the intervention effect, and intervention-independent tests (including unstandardized as well as standardized tests), and also the difference in effect measured using standardized and unstandardized tests (including the self-developed tests). Both distinctions were relevant, but the latter explained by far the most variance in effect. In accordance with the findings of Chiu (1998), we found that standardized tests yielded a lower effect than unstandardized ones. This attribute explained the most variance of all attributes examined in this meta-analysis. The difference in effect supports Rosenshine's (1994) suggestion that standardized tests are less sensitive in measuring intervention effects than other types of tests. Standardized tests probably measure student performance in a broad sense and are not focused on task-specific student performances. The interventions included in this meta-analysis may have been focused on a relatively small domain and, therefore, did not improve student performance as much on standardized tests as on other tests.

Limitations

The interventions examined in this study differed from one another in many respects. However, we were not able to take all these differences into account. To give an example, a meta-analysis can include one variable per approximately 10 interventions (Borenstein et al., 2009). Although with 95 interventions in total, we were able to include many attributes that had been coded, we could not control for all. We therefore chose to only include the learning strategies that appeared significant in a prior metastudy as covariates (Donker et al., 2014) and then to analyze the effects of their attributes. Because of this constraint in sample size, we do not know if a variable that was not included in the analysis might have influenced the results. Furthermore, all interventions were focused on learning strategy instruction; therefore, we do not know for sure whether the findings on the influence of the interventions' attributes can be generalized to other types of educational instruction. However, we believe that although the effect sizes of the attributes of the various types of educational interventions may differ to some extent, the trends found can also be related to fields other than learning strategy instruction. Another aspect that most likely confused our findings is that the effect sizes of the interventions were measured using a large spectrum of different tests. These tests were, of course, not calibrated, whereby the ways in which student performance was measured varied. In addition, on taking a closer look at the effect sizes estimated, we noticed that there were large differences within individual interventions. This large differentiation made it more difficult to compare the effects of the interventions and analyze the impact of the attributes.

Nevertheless, despite the limitations of meta-analytical research, this approach yields more reliable conclusions than a single primary study does. In meta-analysis, the measurement and implementation errors of the primary studies are averaged out, which leads to more balanced results. Therefore, it is a valuable method to summarize the findings of primary studies. As such, the findings of the current study contribute to the body of knowledge about the influence of the attributes of educational interventions and their effect studies on the effectiveness of these interventions and the quality of their effect studies, in particular that of learning strategy instruction.

Practical Implications

The four attributes that appear likely to be most important together explained 63.2% of the variance in intervention effect size. This finding implies that the effectiveness of interventions depends to a large degree on how they are implemented and how their effects are examined. This conclusion stresses the importance of implementing and testing an intervention with care. It also makes clear that one should be cautious when making statements about interventions. The results of the current study show that the effect cannot automatically be ascribed to the instructional focus of the intervention. The other attributes related to its implementation and methods used to evaluate its effects also have to be considered.

The findings show that interventions implemented by the assistant or the researcher have the highest impact; either because these implementers are more inclined to "teach to the test," or because they are highly motivated and have the confidence to make the intervention a success. If the latter explanation holds true, this is a strong indication for researchers to take teacher motivation into account when asking teachers to implement their intervention. It would be wise to first focus on creating a broad basis of support among the teachers for the intervention. Furthermore, researchers should provide teachers with thorough instructions on how to implement the intervention properly and coach them during the implementation phase. Although teachers do not always look forward to having a visitor in their classroom to watch their teaching, they do see the benefits of it, as Datnow and Castellano (2000) described in their study of the implementation of a new teaching method. Provision of good and solid instruction to teachers could be especially important from a long-term point of view, because once the teachers have internalized this new knowledge, they can continue to work with it even when the research experiment has ended. As well as enhancing students' performance in the experimental groups, this way of working may also positively influence the achievement levels of students in the future. Therefore, we recommend that researchers prepare teachers thoroughly before they have them implement an intervention. It may also help to let teachers co-construct the intervention with the researcher. In this way, the intervention can be adapted to fit the specific school (and teacher) and still maintain its effective components (Datnow & Stringfield, 2000). At first sight, these recommendations may appear to be merely aspects that increase the costs of an experiment. However, if research indicates that it works, it may be relevant to consider whether the benefits in (long-term) student performance are worth the extra costs (Creemers & van der Werf, 2000).

The findings of this meta-analysis also showed that allowing cooperation between students was not effective in enhancing student performance. This might be because cooperation was merely part of an intervention that was focused on something else. Other research that was focused on cooperative learning did find a positive effect of this teaching method. We therefore recommend that if cooperative learning is used as a teaching method to facilitate the learning of other aspects, it should be implemented with great care.

Furthermore, in line with Rosenshine (1994), we strongly recommend that future researchers add standardized tests to their set of measurement instruments (in addition to any self-developed or other tests). Although these tests might be less sensitive in measuring intervention effects, they have the great advantage that they facilitate a more reliable comparison of multiple intervention effects than do unstandardized measurements.

Because the duration of an intervention was found to have no impact on the effect of an educational intervention, we conclude that it is more cost-effective to reduce the number of sessions in the intervention trajectory. However, we also believe that more research is needed to determine a proper threshold. The findings of the current study indicate that the duration of each session should not be too short, as longer sessions had a very small positive effect on the intervention's success.

Future Research Directions

Our results indicated that the duration of the interventions had no significant influence on their effectiveness. It is not unlikely, though, that longer interventions have more impact on the *overall* performance of students. Longer interventions may be less task-specific than short ones, because they are more apt to provide students with opportunities to transfer their learning to other, similar tasks. We checked this hypothesis by examining the effect of duration after taking the type of measurement instrument into account. We assumed that standardized tests did not focus on task-specific student performance, whereas unstandardized tests would do so to a greater extent. However, we still found no effect. Either the standardized tests were still quite task-specific, or the duration of the intervention simply plays no role whatsoever. Nevertheless, we do recommend a further examination of the relationship between overall student performance and the duration of interventions.

If short interventions also yield an intervention effect, it would be a waste of time and money to invest in longer interventions. An example of a good analysis in this context would be to estimate intervention effects using tests to monitor student progress. In the Netherlands, for instance, primary schools use a system of standardized and method-independent tests, which measures student performance in several subject domains. These tests are taken one or more times per year. In this way, the progress in students' performance is well monitored. It would be interesting to see if these interventions indeed improve students' future scores on such tests, and if their duration has any effect on the results. A final recommendation for future meta-analytical research is to again analyze the influence of cooperative learning as a supportive teaching method on the effectiveness of other types of educational interventions. Until now, the meta-analyses that included this element have revealed mixed results.

References

Studies included in the meta-analysis are marked with an asterisk.

Abrami, P. C., Poulsen, C., & Chambers, B. (2004). Teacher motivation to implement an educational innovation: Factors differentiating users and non-users of cooperative learning. *Educational Psychology*, *24*, 201–216. doi:10.1080/0144341032000160146

*Aleven, V. A. W.M. M., & Koedinger, K. R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based cognitive tutor. *Cognitive Science: A Multidisciplinary Journal*, *26*, 147–179. doi:10.1207/s15516709cog2602_1

*Allen, K. D., & Hancock, T. E. (2008). Reading comprehension improvement with individualized cognitive profiles and metacognition. *Literacy Research and Instruction*, *47*, 124–139. doi:10.1080/19388070801938320

Bangert-Drowns, R. L., Hurley, M. M., & Wilkinson, B. (2004). The effects of school-based writing-to-learn interventions on academic achievement: A meta-analysis. *Review of Educational Research*, *74*, 29–58. doi:10.3102/00346543074001029

*Blank, L. M. (2000). A metacognitive learning cycle: A better warranty for student understanding? *Science Education*, *84*, 486–506. doi:10.1002/1098-237X(200007)84:4<486::AID-SCE4>3.0.CO;2-U

Boekaerts, M. (1997). Self-regulated learning: A new concept embraced by researchers, policy makers, educators, teachers, and students. *Learning and Instruction*, *7*, 161–186. doi:10.1016/S0959-4752(96)00015-1

Boekaerts, M., Pintrich, P. R., & Zeidner, M. (2000). *Handbook of self-regulation*. San Diego, CA: Academic Press.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, England: Wiley. doi:10.1002/9780470743386

*Boulware-Gooden, R., Carreker, S., Thornhill, A., & Malatesha, J. R. (2007). Instruction of metacognitive strategies enhances reading comprehension and vocabulary achievement of third-grade students. *The Reading Teacher*, *61*, 70–77. doi:10.1598/RT.61.1.7

*Bruce, M. E., & Robinson, G. L. (2001, July). *The clever kid's reading program: Metacognition and reciprocal teaching*. Paper presented at the Annual European Conference on Reading, Dublin, Ireland.

*Brunstein, J. C., & Glaser, C. (2011). Testing a path-analytical mediation model of how self-regulated writing skills improve fourth-graders' composition skills: A randomized controlled trial. *Journal of Educational Psychology*, *103*, 922–938. doi:10.1037/a0024622

*Camahalan, F. M. G. (2006). Effects of self-regulated learning on mathematics achievement of selected Southeast Asian children. *Journal of Instructional Psychology*, *33*, 194–205.

*Cantrell, S. C., Almasi, J. F., Carter, J. C., Rintamaa, M., & Madden, A. (2010). The impact of a strategy-based intervention on the comprehension and strategy use of struggling adolescent readers. *Journal of Educational Psychology*, *102*, 257–280. doi:10.1037/a0018212

Chiu, C. W. T. (1998, April). *Synthesizing metacognitive interventions: What training characteristics can improve reading performance?* Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Creemers, B. P. M. (1994). *The effective classroom*. London, England: Cassell.

Creemers, B. P. M., & van der Werf, M. P. C. (2000). Economic viewpoints in educational effectiveness: Cost-effectiveness analysis of an educational improvement project. *School Effectiveness and School Improvement*, *11*, 361–384. doi:10.1076/0924-3453(200009)11:3;1-G;FT361

Datnow, A., & Castellano, M. (2000). Teachers' responses to success for all: How beliefs, experiences and adaptations shape implementation. *American Educational Research Journal*, *37*, 775–799. doi:10.3102/00028312037003775

Datnow, A., & Stringfield, S. (2000). Working together for reliable school reform. *Journal of Education for Students Placed at Risk*, *5*, 183–204. doi:10.1080/10824669.2000.9671386

*De Acedo Lizarraga, M. L. S., de Acedo Baquedano, M. T. S., & Oliver, M. S. (2010). Stimulation of thinking skills in high school students. *Educational Studies*, *36*, 329–340. doi:10.1080/03055690903425003

*Dejonckheere, P., Van de Keere, K., & Tallir, I. (2011). Are fourth and fifth grade children better scientists through metacognitive learning? *Electronic Journal of Research in Educational Psychology*, *9*, 133–156.

Dignath, C., & Büttner, G. (2008). Components of fostering self-regulated learning among students: A meta-analysis on intervention studies at primary and secondary school level. *Metacognition Learning*, *3*, 231–264. doi:10.1007/s11409-008-9029-x

Donker, A. S., de Boer, H., Kostons, D., Dignath-van Ewijk, C. C., & van der Werf, M. P. C. (2014). Effectiveness of learning strategy instruction on academic performance: A meta-analysis. *Educational Research Review*, *11*, 1–26. doi:10.1016/j.edurev.2013.11.002

*Dresel, M., & Haugwitz, M. (2008). A computer-based approach to fostering motivation and self-regulated learning. *Journal of Experimental Education*, *77*, 3–18. doi:10.3200/JEXE.77.1.3-20

Dynarski, M., Agodini, R., Heaviside, S., Novak, T., Carey, N., Campuzano, L., . . .Sussex, W. (2007). *Effectiveness of reading and mathematics software products: Findings from the first student cohort*. Washington, DC: U.S. Department of Education, Institute of Education Sciences.

*Erktin, E. (2004). Teaching thinking for mathematics through the enhancement of metacognitive skills. *Research in the Schools*, *11*, 3–13.

*García-Sánchez, J. N., & Fidalgo-Redondo, R. (2006). Effects of two types of self-regulatory instruction programs on students learning disabilities in writing products, processes, and self-efficacy. *Learning Disability Quarterly*, *29*, 181–211. doi:10.2307/30035506

*Glaser, C., & Brunstein, J. C. (2007). Improving fourth-grade students' composition skills: Effects of strategy instruction and self-regulation procedures. *Journal of Educational Psychology*, *99*, 297–310. doi:10.1037/0022-0663.99.2.297

*Graham, S., Harris, K. R., & Mason, L. (2005). Improving the writing performance, knowledge, and self-efficacy of struggling young writers: The effects of self-regulated strategy development. *Contemporary Educational Psychology*, *30*, 207–241. doi:10.1016/j.cedpsych.2004.08.001

Guskey, T. R. (2007). Closing achievement gaps: Revisiting Benjamin S. Bloom's "learning for mastery." *Journal of Advanced Academics*, *19*, 8–31.

Guskey, T. R., & Pigott, T. D. (1988). Research on group-based mastery learning programs: A meta-analysis. *Journal of Educational Research*, *81*, 197–216.

*Guterman, E. (2003). Integrating written metacognitive awareness guidance as a "psychological tool" to improve student performance. *Learning and Instruction*, *13*, 633–651. doi:10.1016/S0959-4752(02)00070-1

*Harris, K. R., Graham, S., & Mason, L. H. (2006). Improving the writing, knowledge, and motivation of struggling young writers: Effects of self-regulated strategy development with and without peer support. *American Educational Research Journal*, *43*, 295–340. doi:10.3102/00028312043002295

Hattie, J., Biggs, J., & Purdie, N. (1996). Effects of learning skills interventions on student learning: A meta-analysis. *Review of Educational Research*, *66*, 99–136. doi:10.3102/00346543066002099

*Hauptman, H., & Cohen, A. (2011). The synergetic effect of learning styles on the interaction between virtual environments and the enhancement of spatial thinking. *Computers & Education*, *57*, 2106–2117. doi:10.1016/j.compedu.2011.05.008

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

*Huff, J. D., & Nietfeld, J. L. (2009). Using strategy instruction and confidence judgments to improve metacognitive monitoring. *Metacognition and Learning*, *4*, 161–176. doi:10.1007/s11409-009-9042-8

*Jacobse, A. E., & Harskamp, E. (2009). Student-controlled metacognitive training for solving word problems in primary school mathematics. *Educational Research and Evaluation*, *15*, 447–463. doi:10.1080/13803610903444519

*Kaniel, S., Licht, P., & Peled, B. (2000). The influence of metacognitive instruction of reading and writing strategies on positive transfer. *Gifted Education International*, *51*, 45–63. doi:10.1177/026142940001500106

*Kapa, E. (2001). A metacognitive support during the process of problem solving in a computerized environment. *Educational Studies in Mathematics*, *47*, 317–336. doi:10.1023/A:1015124013119

*Kapa, E. (2007). Transfer from structured to open-ended problem solving in a computerized metacognitive environment. *Learning and Instruction*, *17*, 688–707. doi:10.1016/j.learninstruc.2007.09.019

*Kim, H. J., & Pedersen, S. (2011). Advancing young adolescents' hypothesis-development performance in a computer-supported and problem-based learning environment. *Computers and Education*, *57*, 1780–1789. doi:10.1016/j.compedu.2011.03.014

*Kramarski, B., & Dudai, V. (2009). Group-metacognitive support for online inquiry in mathematics with differential self-questioning. *Journal of Educational Computer Research*, *40*, 377–404. doi:10.2190/EC.40.4.a

*Kramarski, B., & Gutman, M. (2006). How can self-regulated learning be supported in mathematical e-learning environments? *Journal of Computer Assisted Learning*, *22*, 24–33. doi:10.1111/j.1365-2729.2006.00157.x

*Kramarski, B., & Hirsch, C. (2003). Effects of computer algebra system (CAS) with metacognitive training on mathematical reasoning. *Educational Media International*, *40*, 249–257. doi:10.1080/0952398032000113167

*Kramarski, B., & Mevarech, Z. R. (2003). Enhancing mathematical reasoning in classroom: The effects of cooperative learning and the metacognitive training. *American Educational Research Journal*, *40*, 281–310. doi:10.3102/00028312040001281

*Kramarski, B., Mevarech, Z. R., & Arami, M. (2002). The effects of metacognitive instruction on solving mathematical authentic tasks. *Educational Studies in Mathematics*, *49*, 225–250. doi:10.1023/A:1016282811724

*Kramarski, B., Mevarech, Z. R., & Lieberman, A. (2001). Effects of multilevel versus unilevel metacognitive training on mathematical reasoning. *Journal of Educational Research*, *94*, 292–300. doi:10.1080/00220670109598765

*Kramarski, B., & Mizrachi, N. (2006). Online discussion and self-regulated learning: Effects of instructional methods on mathematical literacy. *Journal of Educational Research*, *99*, 218–230. doi:10.3200/JOER.99.4.218-231

*Kramarski, B., & Ritkof, R. (2002). The effects of metacognition and email interactions on learning graphing. *Journal of Computer Assisted Learning*, *18*, 33–43. doi:10.1046/j.0266-4909.2001.00205.x

*Kramarski, B., & Zoldan, S. (2008). Using errors as springboards for enhancing mathematical reasoning with three metacognitive approaches. *Journal of Educational Research*, *102*, 137–151. doi:10.3200/JOER.102.2.137-151

Lewis, M. (2007, February). *Stepwise versus hierarchical regression: Pros and cons*. Paper presented at the annual meeting of the Southwest Educational Research Association, San Antonio, TX.

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

*Lubliner, S., & Smetana, L. (2005). The effects of comprehensive vocabulary instruction on Title I students' metacognitive word-learning skills and reading comprehension. *Journal of Literacy Research*, *37*, 163–200. doi:10.1207/s15548430jlr3702_3

*Mason, L. H. (2004). Explicit self-regulated strategy development versus reciprocal questioning: Effects on expository reading comprehension among struggling readers. *Journal of Educational Psychology*, *96*, 238–296. doi:10.1037/0022-0663.96.2.283

Mayer, R. E. (2008). *Learning and instruction* (2nd ed.). Upper Saddle River, NJ: Pearson Merrill Prentice Hall.

*Mevarech, Z. R., & Kramarski, B. (2004). The effects of metacognitive training versus worked-out examples on students' mathematical reasoning. *British Journal of Educational Psychology*, *73*, 449–471. doi:10.1348/000709903322591181

*Meyer, E., Abrami, P. C., Wade, C. A., Aslan, O., & Deault, L. (2010). Improving literacy and metacognition with electronic portfolio's: Teaching and learning with ePEARL. *Computers & Education*, *55*, 84–91. doi:10.1016/j.compedu.2009.12.005

*Michalsky, T., Mevarech, Z. R., & Haibi, L. (2009). Elementary school children reading scientific texts: Effects of metacognitive instruction. *Journal of Educational Research*, *102*, 363–376. doi:10.3200/JOER.102.5.363-376

*Molenaar, I., Chiu, M. M., & Sleegers, P. (2011). Scaffolding of small-groups' metacognitive activities with an avatar. *International Journal of Computer-Supported Collaborative Learning*, *6*, 601–624. doi:10.1007/s11412-011-9130-z

*Mourad, A. E. (2009). The effectiveness of a program based on self-regulated strategy development on the writing skills of writing-disabled secondary school students. *Electronic Journal of Research in Educational Psychology*, *7*, 5–24.

Orwin, R. G. (1983). A fail-safe *N* for effect size in meta-analysis. *Journal of Educational Statistics*, *8*, 157-159. doi:10.2307/1164923

*Pennequin, V., Sorel, O., Nanty, I., & Fontaine, R. (2010). Metacognition and low achievement in mathematics: The effect of training in the use of metacognitive skills to solve mathematical word problems. *Thinking and reasoning*, *16*, 198–220. doi: 10.1080/13546783.2010.509052

*Perels, F., Dignath, C., & Schmitz, B. (2009). Is it possible to improve mathematical achievement by means of self-regulation strategies? Evaluation of an intervention in regular math classes. *European Journal of Psychology of Education*, *24*, 17–31. doi:10.1007/BF03173472

*Perels, F., Gurtler, T., & Schmitz, B. (2005). Training of self-regulatory and problem-solving competence. *Learning and Instruction*, *15*, 123–139. doi:10.1016/j.learninstruc.2005.04.010

*Peters, E. E., & Kitsantas, A. (2010a). Self-regulation of student epistemic thinking in science: The role of metacognitive prompts. *Educational Psychology*, *30*, 27–52. doi:10.1080/01443410903353294

*Peters, E. E., & Kitsantas, A. (2010b). The effect of nature of science metacognitive prompts on science students' content and nature of science knowledge, metacognition, and self-regulatory efficacy. *School Science and Mathematics*, *110*, 382–396. doi:10.1111/j.1949-8594.2010.00050.x

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2007). Performance of the trim and fill method in the presence of publication bias and between-study heterogeneity. *Statistics in Medicine*, *26*, 4544–4562. doi:10.1002/sim.2889

Pressley, M. (2002). Comprehension strategies instruction: A turn-of-the-century status report. In C. C. Block & M. Pressley (Eds.), *Comprehension instruction: Research-based best practices* (pp. 11–27). New York, NY: Guilford Press.

Raudenbush, S. W., Bryk, A. S., & Congdon, R. (2004). HLM 6 for Windows [Computer software]. Skokie, IL: Scientific Software International.

*Reynolds, G. A., & Perin, D. (2009). A comparison of text structure and self-regulated writing strategies for composing from sources by middle school students. *Reading Psychology*, *30*, 265–300. doi:10.1080/02702710802411547

Rosenshine, B. (1994). The conduct and reporting of intervention studies. *Educational Psychology Review*, *6*, 245–254. doi:10.1007/BF02213186

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*, 638–641. doi:10.1037//0033-2909.86.3.638

Slavin, R. E. (1991). Synthesis of research on cooperative learning. *Educational Leadership*, *48*(5), 71–82.

Slavin, R. E., & Lake, C. (2009, March). *Effective educational programs: Meta-findings from Best Evidence Encyclopedia*. Paper presented at the annual meetings of the Society for Research on Educational Effectiveness, Crystal City, VA.

Slavin, R., & Madden, N. A. (2011). Measures inherent to treatments in program effectiveness reviews. *Journal of Research on Educational Effectiveness*, *4*, 370–380. doi:10.1080/19345747.2011.558986

*Souvignier, E., & Mokhlesgerami, J. (2006). Using self-regulation as a framework for implementing strategy instruction to foster reading comprehension. *Learning and Instruction*, *16*, 57–71. doi:10.1016/j.learninstruc.2005.12.006

*Stoeger, H., & Ziegler, A. (2008). Evaluation of a classroom based training to improve self-regulation in time management tasks during homework activities with fourth graders. *Metacognition and Learning*, *3*, 207–230. doi:10.1007/s11409-008-9027-z

*Stoeger, H., & Ziegler, A. (2010). Do pupils with differing cognitive abilities benefit similarly from a self-regulated learning training program? *Gifted Education International*, *26*, 110–123. doi:10.1177/026142941002600113

*Tajika, H., Nakatsu, N., Nozaki, H., Neumann, E., & Maruno, S. (2007). Effects of self-explanation as a metacognitive strategy for solving mathematical word problems. *Japanese Psychological Research*, *49*, 222–233. doi:10.1111/j.1468-5884.2007.00349.x

*Talebi, S. H. (2009). An attempt towards learner autonomy in L2 (English) and L3 (Arabic) reading through cognitive and metacognitive reading strategy instruction (CMRSI) in L2. *Indian Journal of Applied Linguistics*, *35*, 101–112.

*Teong, S. K. (2003). The effect of metacognitive training on mathematical word-problem solving. *Journal of Computer Assisted Learning*, *19*, 46–55. doi:10.1046/j.0266-4909.2003.00005.x

Thompson, B. (1995). Stepwise regression and stepwise discriminant analysis need not apply here: A guidelines editorial. *Educational and Psychological Measurement*, *55*, 525-534. doi:10.1177/0013164495055004001

*Torrance, M., Fidalgo, R., & García, J-N. (2007). The teachability and effectiveness of cognitive self-regulation in sixth-grade writers. *Learning and Instruction*, *17*, 265–285. doi:10.1016/j.learninstruc.2007.02.003

*Tracy, B., Reid, R., & Graham, S. (2009). Teaching young students strategies for planning and drafting stories: The impact of self-regulated strategy development. *Journal of Educational Research*, *102*, 323–331. doi:10.3200/JOER.102.5.323-332

*Van Keer, H., & Vanderlinde, R. (2010). The impact of cross-age peer tutoring on third and sixth graders' reading strategy awareness, reading strategy use, and reading comprehension. *Middle Grades Research Journal*, *5*, 33–45.

*Vaughn, S., Klingner, J. K., & Swanson, E. A. (2011). Efficacy of collaborative strategic reading with middle school students. *American Educational Research Journal*, *48*, 938–964. doi:10.3102/0002831211410305

Wehby, J. H., Maggin, D. M., Johnson, L., & Symons, F. J. (2010, March). *Improving intervention implementation and fidelity in evidence-based practice: Integrating teacher preference into intervention selection*. Paper presented at the annual meetings of the Society for Research on Educational Effectiveness, Washington, DC.

Weinstein, C. E., & Mayer, R. E. (1986). The teaching of learning strategies. In M. Wittrock (Ed.), *Handbook of research on teaching* (pp. 315–327). New York, NY: Macmillan.

*Wright, J., & Jacobs, B. (2003). Teaching phonological awareness and metacognitive strategies to children with reading difficulties: A comparison of two instructional methods. *Educational Psychology*, *23*, 17–47. doi:10.1080/01443410303217

Zimmerman, B. J. (2002). Becoming a self-regulated learner: an overview. *Theory Into Practice*, *41*, 64–70. doi:10.1207/s15430421tip4102_2

*Zion, M., Michalsky, T., & Mevarech, Z. R. (2005). The effects of metacognitive instruction embedded within an asynchronous learning network on scientific inquiry skills. *International Journal of Science Education*, *27*, 957–983. doi:10.1080/09500690500068626

Authors

HESTER DE BOER is a postdoctoral researcher at the Institute for Educational Research of the University of Groningen, Grote Rozenstraat 3, 9712 TG Groningen, Netherlands; e-mail: *hester.de.boer@rug.nl*. Her research interests include self-regulated learning and expectations and aspirations of teachers and parents. Her methodological expertise is in meta-analysis.

ANOUK S. DONKER is a doctoral student at the Institute for Educational Research of the University of Groningen, Grote Rozenstraat 3, 9712 TG Groningen, Netherlands; e-mail: *a.s.donker-bergstra@rug.nl*. Her research interests include metacognition, self-regulated learning, and comprehensive reading.

MARGARETHA P. C. VAN DER WERF is full professor of learning and instruction at the University of Groningen, Grote Rozenstraat 3, 9712 TG Groningen, Netherlands; e-mail: *m.p.c.van.der.werf@rug.nl*. Her main interests include educational effectiveness, civics and citizenship education, and the influence of psychological precursors of school success. Her expertise is in conducting large-scale multilevel longitudinal research as well as evidence-based field experiments.