

A Meta-Analysis of the Effects of Classroom Management Strategies and Classroom Management Programs on Students' Academic, Behavioral, Emotional, and Motivational Outcomes

Hanke Korpershoek
Truus Harms
Hester de Boer
Mechteld van Kuijk
Simone Doolaard
University of Groningen

This meta-analysis examined which classroom management strategies and programs enhanced students' academic, behavioral, social-emotional, and motivational outcomes in primary education. The analysis included 54 random and nonrandom controlled intervention studies published in the past decade (2003–2013). Results showed small but significant effects (average $g = 0.22$) on all outcomes, except for motivational outcomes. Programs were coded for the presence/absence of four categories of strategies: focusing on the teacher; on student behavior; on students' social-emotional development, and on teacher–student relationships. Focusing on the students' social-emotional development appeared to have the largest contribution to the interventions' effectiveness, in particular on the social-emotional outcomes. Moreover, we found a tentative result that students' academic outcomes benefited from teacher-focused programs.

KEYWORDS: review, meta-analysis, classroom management strategies, classroom management programs, student outcomes

Effective education refers to the degree to which schools are successful in accomplishing their educational objectives. The findings of numerous studies have shown that teachers play a key role in shaping effective education (Hattie, 2009). The differences in achievement between students who spend a year in a class with a highly effective teacher as opposed to a highly ineffective teacher are startling. Effective teaching and learning cannot take place in poorly managed classrooms (V. F. Jones & Jones, 2012; Marzano, Marzano, & Pickering, 2003;

Van de Grift, Van der Wal, & Torenbeek, 2011). The main objective of the present study was therefore to conduct a meta-analysis of the effects of various classroom management strategies (CMS) and classroom management programs (CMP) aimed at improving students' behavior and enhancing their academic performance in primary education.

Effective CMS support and facilitate effective teaching and learning. Effective classroom management is generally based on the principle of establishing a positive classroom environment encompassing effective teacher–student relationships (Wubbels, Brekelmans, Van Tartwijk, & Admiraal, 1999). Effective CMS focus more strongly on preventive rather than reactive classroom management procedures (Lewis & Sugai, 1999). An example of a widely used—and generally effective—preventive strategy among teachers in primary education is that classroom rules are negotiated instead of imposed (Marzano et al., 2003). Teachers, however, also frequently use reactive strategies (e.g., punishing disruptive students; Rydell & Henricsson, 2004; Shook, 2012), although it is unclear whether these strategies effectively change student behavior.

The frequent use of reactive strategies may be caused by a lack of knowledge about the effectiveness of preventive strategies (e.g., J. H. Peters, 2012) or by a lack of belief in their effectiveness (e.g., Smart & Brent, 2010). One example is that student teachers are generally advised to be as strict as possible in the first week of their internship and then slowly to become less authoritarian, although first establishing positive teacher–student relationships has been proven far more effective in regulating student behavior (e.g., Bohn, Roehrig, & Pressley, 2004). When teachers feel uncertain about using preventive strategies (see O'Neill & Stephenson, 2012), such as negotiating about classroom rules, they often keep using the presumably less effective reactive strategies (Rydell & Henricsson, 2004; Woodcock & Reupert, 2012).

Daily practice in education has changed rapidly. It is increasingly characterized by student-centered approaches to learning as opposed to teacher-centered, with a large emphasis on students' metacognitive skills (e.g., self-regulated learning strategies; Dignath, Buettner, & Langfeldt, 2008) and cooperative learning (e.g., Kagan, 2005; Wubbels, Den Brok, Veldman, & Van Tartwijk, 2006). Moreover, more and more technology is finding its way into classrooms, through the use of interactive whiteboards, tablets, or laptops (Schussler, Poole, Whitlock, & Evertson, 2007). These changes have had a large impact on the demands placed on teachers' classroom management skills (e.g., rules and procedures to facilitate cooperative learning). Although, to the best of our knowledge, no studies have been conducted to explicitly compare the effectiveness of particular CMS in more traditional versus more modern classrooms, an up-to-date overview of studies conducted in the past decade is expected to provide insight into which CMS have been proven (still) to be effective in modern classrooms.

Definition of Classroom Management

Evertson and Weinstein (2006) referred in their definition of classroom management to the actions teachers take to create a supportive environment for the academic and social-emotional learning of students. They described five types of actions. To attain a high quality of classroom management, teachers

must (a) develop caring, supportive relationships with and among students (see also Marzano et al., 2003); (b) organize and implement instruction in ways that optimize students' access to learning; (c) encourage students' engagement in academic tasks, which can be done by using group management methods (e.g., by establishing rules and classroom procedures, see Marzano et al., 2003); (d) promote the development of students' social skills and self-regulation, which Marzano et al. (2003) referred to as making students responsible for their behavior; and (e) use appropriate interventions to assist students with behavior problems.

The last two actions proposed by Evertson and Weinstein (2006) indicate that effective classroom management improves student behavior. Hence, classroom management is an ongoing interaction between teachers and their students. Brophy (2006) presented a similar definition: "Classroom management refers to actions taken to create and maintain a learning environment conducive to successful instruction (arranging the physical environment, establishing rules and procedures, maintaining students' attention to lessons and engagement in activities)" (p. 17). Both definitions emphasize the importance of actions taken by the teacher to facilitate learning among the students.

CMS and Different Classifications of CMS

As stated above, classroom management is about creating inviting and appealing environments for student learning. CMS are tools that the teachers can use to help create such an environment, ranging from activities to improve teacher–student relationships to rules to regulate student behavior. Only when the efforts of management fail should teachers have to resort to reactive, controlling strategies. Therefore, it is important to distinguish between preventive and reactive CMS (see also Lane, Menzies, Bruhn, & Crnobori, 2011). For example, the establishment of rules and procedures and favorable teacher–student relationships are considered preventive strategies, whereas disciplinary interventions such as giving warnings or punishments are considered reactive strategies. In a similar vein, Froyen and Iverson (1999) used the concepts management of content (e.g., space, materials, equipment, movement, and lessons) and management of covenant (e.g., social dynamics and interpersonal relationships) for preventive strategies and management of conduct (e.g., disciplinary problems) for reactive strategies when referring to classroom management.

A separate group of CMS are group contingencies, which represent various reinforcement strategies aimed at improving student behavior or performance. These include preventive and reactive strategies. These group contingencies can be classified into three types as discussed in Kelshaw-Levering, Sterling-Turner, and Henry (2000): independent, interdependent, and dependent group contingencies. Independent group contingencies refer to reinforcement interventions that apply the same assessment criteria and reinforcements to each child (e.g., all children should pass the same swimming test before they get a diploma). Dependent group contingencies, on the other hand, refer to interventions that require a single student (or a few students) to reach a designated criterion for the whole group to receive reinforcement (e.g., when a student attains 100% on a test, the teacher will hand out sweets to the entire class). Interdependent group contingencies require

the whole student group to reach a designated criterion in order to receive reinforcement (e.g., group members need to collaborate on a team project and the entire team receives a grade for their end product).

When comparing the above-mentioned classifications of CMS (preventive/reactive, management of content/covenant/conduct), we did not find a systematic classification of classroom management interventions that covers the whole range of classroom management dimensions based on Evertson and Weinstein's (2006) definition of classroom management, the most exhaustive description of what classroom management entails from our perspective. Improving student behavior (e.g., self-control) is an important goal in many CMP nowadays, although this student component is underrepresented in the different classifications mentioned above. Moreover, in many interventions, both preventive and reactive strategies are used. Therefore, we propose the following classification of classroom management interventions, based on their primary focus:

1. *Teachers' behavior-focused interventions*: The focus of the intervention is on improving teachers' classroom management (e.g., keeping order, introducing rules and procedures, disciplinary interventions) and thus on changing the teachers' behavior. This type is a representation of group management methods (Evertson & Weinstein, 2006). Both preventive and reactive interventions are included in this category.
2. *Teacher–student relationship–focused interventions*: The focus of the intervention is on improving the interaction *between* teachers and students, that is, on developing caring, supportive relationships. Only preventive interventions are included in this category. This type is a representation of the supportive teacher–student relation (Evertson & Weinstein, 2006). Interventions focusing on relations between students *only* are not included here.
3. *Students' behavior-focused interventions*: The focus of the intervention is on improving student behavior, for example, via group contingencies or by improving self-control among all students. Both preventive and reactive interventions are included in this category. This type is a representation of the students' self-regulation (Evertson & Weinstein, 2006) or what Marzano et al. (2003) referred to as students' responsibility for their own behavior.
4. *Students' social-emotional development-focused interventions*: The focus of the intervention is on improving students' social-emotional development, such as enhancing their feelings of empathy for other children. Both preventive and reactive interventions are included in this category. This type is a representation of the students' social skills (Evertson & Weinstein, 2006).

Some CMP may fit into more than one of these categories as the types are not mutually exclusive. The proposed classification was used in the meta-analysis to identify the differential effects of different types of interventions. Moreover, it is possible that broader interventions that have multiple foci may result in stronger

effects than interventions that have one primary focus, or that a particular combination of foci may be more effective than other combinations.

Prior Meta-Analyses

Three relevant prior meta-analyses are summarized in this section. The study by Marzano et al. (2003) is the most recent meta-analysis of effective classroom management, based on 101 studies published between 1967 and 1997. The participants were primary and secondary school students in regular as well as in special education. About half of the studies were based on a single participant, the other half on groups of students. Marzano et al. studied several components of teachers' effective classroom management. Based on 10 studies, the researchers reported an effect size that is clinically and statistically significant, $d = -0.76$, confidence interval [CI; $-0.93, -0.60$], for rules and procedures. Their results can be interpreted as follows: In classrooms focused on effective use of rules and procedures, the average number of classroom interruptions was 0.76 standard deviations less than in classrooms that were not focused on these techniques. For disciplinary interventions, the effect size was $d = -0.91$ (CI not reported); for teacher–student relationships, $d = -0.87$, CI [$-1.00, -0.74$]; for mental set, $d = -1.29$, CI [$-1.49, -1.10$]; and for student responsibility, $d = -0.69$, CI [$-0.83, -0.56$], based on 68, 4, 5, and 28 studies, respectively. The meta-analysis included seven studies in which the effects of CMS on engagement were measured and five studies in which the effects on achievement were measured; the results revealed average effects of 0.62 and 0.52 standard deviations higher, respectively.

A limitation of Marzano et al.'s (2003) meta-analysis is that the authors did not report how they performed the literature search (i.e., what search terms and eligibility criteria were used) and how the meta-analysis was executed. As a result, the exact methods used to arrive at their findings are not known. For example, it is unclear (a) how the authors selected the studies, (b) whether the primary studies were experiments in which the effects of CMS were examined rather than correlational studies, and (c) whether a control group was always used. Nonetheless, Marzano et al.'s results do suggest that CMS are important for creating an orderly and harmonious learning environment.

In another study, Oliver, Wehby, and Reschly (2011) reported on the effects of universal, whole-class classroom management procedures on problem student behavior. Although the search profile indicates that Oliver et al. (2011) included studies published between 1950 and 2009 on classroom management and classroom organization, the final review included only 12 studies (with only one published after 2000). The participants were both primary and secondary school students, and four studies also included special education classrooms. The findings revealed that teachers' classroom management practices had a significant, positive effect on decreasing problem behavior in the classroom. The researchers reported an effect size of $d = 0.71$, CI [$0.46, 0.96$].

Durlak, Weissberg, Dymnicki, Taylor, and Schellinger (2011) conducted a meta-analysis of 213 school-based, universal (school-wide) social and emotional learning (SEL) programs. These programs are aimed at enhancing students' cognitive, affective, and behavioral competencies such as self-awareness and responsible decision making that lay the foundation for better school adjustment and

academic performance. SEL programs generally include some classroom management components. Focusing on studies that appeared in published or unpublished form before 2007, Durlak et al. selected all school-based universal studies that emphasized the development of one or more SEL skills among students from kindergarten through high school. Hedges's g effect sizes (at the student level) were used, which can be interpreted similarly to Cohen's d effect sizes (Cohen, 1988). They demonstrated that SEL programs significantly improved students' social and emotional skills, $g = 0.57$, CI [0.48, 0.67]. Students who received SEL programs showed more positive social behavior, $g = 0.24$, CI [0.16, 0.32], and had fewer conduct problems, $g = 0.22$, CI [0.16, 0.29]. The effect size for academic achievement was $g = 0.27$, CI [0.15, 0.39]. These effect sizes are slightly lower than the effect sizes reported by Marzano et al. (2003), which may be due to the fact that Durlak et al. reported effect sizes at the student level instead of at the classroom level.

The Present Study

Our main objective was to conduct a meta-analysis of the effects of various CMS/CMP aimed at improving students' behavior and enhancing their academic performance in primary education. In line with Evertson and Weinstein's (2006) definition of classroom management, we focused on the literature on CMS/CMP that support and facilitate both academic and social-emotional learning. As a result, the meta-analysis included studies conducted to examine the effects of CMS/CMP on various student outcomes: (a) academic outcomes (e.g., student performance), (b) behavioral outcomes, (c) social-emotional outcomes, and (d) motivational outcomes. The following research question guided the study: Which CMS and CMP effectively support and facilitate academic, behavioral, social-emotional, and/or motivational outcomes in primary education?

This question was addressed by performing a systematic review of the peer-reviewed classroom management literature published between 2003 and 2013. The present study differs from the previously conducted meta-analyses on classroom management in various ways. An important difference is that we limited the current investigation to whole-class classroom management interventions. Both preventive strategies and reactive strategies can be applied to the entire classroom population (e.g., by discussing classroom rules or giving group detention) or to individual students (e.g., by letting an easily distracted student sit alone during independent seatwork or placing a student temporarily outside the classroom when showing disruptive behavior).

The methods used to investigate strategies to improve individual students' behavior (e.g., students with behavioral and/or emotional disorders) or to discipline individual students (e.g., move seat, isolation time out, detention) are usually single case studies. Without a control group, maturation effects cannot be detected. Particularly for social-emotional and behavioral outcomes, maturation effects are part of students' natural development (e.g., Erikson, 1968). Moreover, it seems that effective management of the whole classroom population (including adequate response to disruptive individual students) is a prerequisite for dealing with students requiring additional behavioral support (see Swinson, Woof, & Melling, 2003).

In prior meta-analyses, the focus was usually broader, and less strict research design criteria were applied. Furthermore, the present meta-analysis builds on the previously conducted meta-analyses by examining recently conducted studies only (i.e., published between 2003 and 2013); thus, the selected studies were conducted in relatively modern classrooms. The results of the meta-analysis therefore give an overview of contemporary CMS/CMP that improve student outcomes. This knowledge base supports teachers in effectively managing their classrooms in current educational settings. Finally, we focused on CMS/CMP interventions that were implemented by teachers in their own classrooms (including school-wide interventions). This limitation is in contrast to some prior publications (e.g., Durlak et al., 2011), in which the interventions were partly implemented by, for example, researchers. For the relevance of our study for educational practice, we sought it more useful to concentrate on interventions implemented by teachers themselves.

Method

Literature Search

The literature search was aimed at identifying studies in which the effectiveness of CMP and their accompanying strategies was investigated. As such, we included the online databases ERIC, Web of Science, PsycINFO, and Picarta from 2003 until 2013. Here, we focused on peer-reviewed journal articles and abstract collections. Although searching for peer-reviewed publications has the disadvantage of neglecting some studies (studies on interventions with no significant effects are less likely to be published at all, and studies described in theses and dissertations are not peer-reviewed), it is a useful criterion for a first selection of studies of sufficient quality. The keyword searches included the following terms: *classroom management, classroom organisation/organization, behavior(al) management, classroom technique(s), teacher/teaching strategy/strategies, classroom discipline, group contingency/contingencies*. These keywords were combined with the following: *academic outcomes, academic achievement, performance, on-task/off-task/time-on-task, student engagement, academic engagement, student behavior, classroom behavior*. Both British English and American English spelling were used. The following wildcards were used: *school**, *contingenc**, *behavio*r**, *teach**. Studies that considered Grades 1 to 6, elementary education, primary education, preschool education, kindergarten, and early childhood education were included. Additionally, the journals *Teaching and Teacher Education* and *Pedagogische Studien* were consulted for relevant references by checking the reference lists of the published articles, as were the publications of Hattie (2009) and Evertson and Weinstein (2006) by checking the reference lists of each chapter.

After the first round, specific classroom management intervention programs were used as additional search terms. The selection of those interventions was based on the results of the first round (the programs that were identified in this round were The Good Behavior Game [GBG], The Color Wheel System, and Classroom Organization and Management Program). Moreover, we found the study by Freiberg and Lapointe (2006), who listed numerous behavioral intervention programs in American education. From this overview, we selected the programs that focused on the entire classroom and used students' behavior or

achievement as outcome measures (the programs identified in this step were the Daily Behavior Report Card, Peacebuilders, Promoting Alternative Thinking Strategies [PATHS], and School-Wide Positive Behavior Support [SWPBS]). Through the Best Evidence Encyclopedia, we found one additional program focused on classroom management. This program, called Consistency Management & Cooperative Discipline, was also included in the additional literature search.

The reference lists of the selected papers were then checked for publications that we had not found in the previous steps. Some of these publications referred to another relevant classroom management intervention program, Zippy's Friends; we decided to use this search term in the databases to find related papers. Finally, we included two new search term combinations, *social-emotional learning* and *social-emotional outcomes* in combination with *school*. This was done because we discovered that some of the interventions we had selected used these terms to explain the content of their intervention (e.g., PATHS). In SEL programs, students "develop skills to recognize and manage their emotions, develop caring and concern for others, make responsible decisions, establish positive relationships, and handle challenging situations effectively" (Weissberg, Resnik, Payton, & O'Brien, 2003, pp. 46–47).

Inclusion Criteria

The studies had to meet the following criteria to be eligible for inclusion: (a) The focus of the study was on CMS of teachers or CMP implemented by teachers in regular, primary school classrooms; (b) The interventions needed to focus on (basically) all students in the classroom, that is, interventions aimed at changing individuals' or small groups' behavior were not eligible; (c) The outcome variable had to include measures of academic outcomes, behavioral outcomes, social-emotional outcomes, motivational outcomes, or other relevant student outcomes (e.g., time-on-task, self-efficacy, peer acceptance); and (d) The studies had to be (quasi-) experimental designs with control groups (no treatment or treatment as usual). They had to meet at least one of the following criteria: (d-1) participants were randomly assigned to treatment and control or comparison conditions; (d-2) participants were matched into treatment and control conditions, and the matching variables included a pretest for the outcome variable or pretest differences were statistically controlled for using ANCOVA; (d-3) if participants were not randomly assigned or matched, the study needed to have a pre–posttest design with sufficient statistical information to derive an effect size or to estimate group equivalence from statistical significance tests.

After the initial screening of the more than 5,000 titles and abstracts to eliminate off-topic papers, 241 studies were selected for further inspection. These studies were divided among three researchers to determine whether they met the inclusion criteria. A second selection round using the four stated criteria was conducted to determine which studies met the inclusion criteria. In this selection round, all studies were initially categorized into three groups: eligible, possibly eligible, not eligible. The researchers met on several occasions to discuss how stringent the inclusion criteria needed to be (e.g., whether kindergarten classrooms are part of primary schools or not. We decided that the studies conducted in these classrooms were eligible for inclusion). Furthermore, all studies that were

initially labeled “possibly eligible” were discussed by the three researchers involved in this selection process. When necessary, a second researcher read the study. Moreover, all studies that were labeled “not eligible” or “eligible” were checked by a second researcher. This second researcher checked the abstract or the full paper when the abstract did not provide enough details. The final decisions for inclusion (“eligible”) were thus based on complete consensus. Following this procedure, 47 studies were selected for the meta-analysis.

The main reasons for excluding 194 studies followed from the inclusion criteria. Most (135 studies; 70%) did not have a suitable research design (no control group, correlational studies, and no empirical data) and therefore did not meet Criterion d. Moreover, 21 studies did not focus on classroom management at all or were not conducted in regular, primary school classrooms (Criterion a). In 10 studies, the intervention was not focused on all students in the classroom (Criterion b), and 7 studies did not include relevant student outcome variables but, for example, included only outcome variables at the school level (e.g., retention rates; Criterion c). For 21 studies, there were other reasons for exclusion: mainly because not enough statistical data were provided to compute effect sizes or the data sets of studies we had already included were used without new relevant additional outcome measures.

Although several school-wide programs focused on antibullying include teacher strategies to reduce problem behavior in class, studies aimed at investigating these programs were excluded from the present study. Several reviews specifically focusing on this topic have already been conducted in recent years (e.g., Ferguson, Miguel, Kilburn, & Sanchez, 2007; Merrell, Gueldner, Ross, & Isava, 2008; Ttofi & Farrington, 2011). For antibullying programs that have been successfully implemented, see Kärnä et al. (2011). Similarly, training programs primarily focused on social skills were excluded because these generally concentrate on enhancing students’ mental resilience rather than their general social-emotional development (e.g., developing empathy). However, when training in social skills was part of another program that met our inclusion criteria, the studies were included in the meta-analysis.

Coding of the Studies

The 47 selected studies were coded for further investigation, initially including the following information: CMS/CMP under investigation (teachers’ behavior focused, teacher–student relationship focused, students’ behavior focused, students’ social-emotional development focused), duration of the intervention, number of intervention sessions, outcome variables (student performance: reading, writing, arithmetic, science, other; time-on-task; student behavior; student engagement), sample characteristics (average students, learning problems, behavioral problems, low socioeconomic status [SES], high SES, grade level, age), country where the study was conducted, educational context (during instruction, independent seatwork, cooperative learning, lesson transitions), classroom setting (group settings, frontal placement, thus facing the teacher), research measurement instrument (designed by authors, unstandardized instrument designed by others, standardized instrument designed by others), design (pre–posttest, control group, random sample), sample size, the reported effects, the number of schools or

classes included, and whether the data were reported at the student, class, or school level.

To code the CMS/CMP under investigation (four categories: teachers' behavior focused, teacher–student relationship focused, students' behavior focused, students' social-emotional development focused), interrater reliabilities were calculated. Two researchers showed 89% agreement (46 studies¹), resulting in an interrater reliability (Cohen's kappa) of .78. The differences in coding concerned 13 studies. In nine cases, one of the researchers had indicated more categories than the other. In these cases, we decided to combine the scores of the two researchers. For the four remaining studies, the coding differences were more substantial. Both researchers reread these articles and changed their initial coding where they thought necessary. This resulted in two studies on which the researchers agreed and two studies in which their scores were combined (as described above).

We were also interested in the effectiveness of frequently used CMP. Therefore, after the initial coding, the studies were categorized into groups with the same intervention (a minimum of three studies per intervention): (a) other, (b) SWPBS, (c) PATHS, (d) GBG, (e) Second Step, and (f) Zippy's Friends. The five programs differed in some respects regarding their main focus.² The SWPBS (Bradshaw, Waasdorp, & Leaf, 2012; Horner et al., 2009) and GBG (Barrish, Saunders, & Wolf, 1969; Flower, McKenna, Bunuan, Muething, & Vega, 2014) programs can be considered both "teachers' behavior focused" and "students' behavior focused," whereas the Fast Track—PATHS (Bierman, Greenberg, & the Conduct Problems Prevention Research Group, 1996; Greenberg & Kusché, 1993, 2002) and Second Step programs (Grossman et al., 1997) can be considered "students' behavior focused" and "students' social-emotional development focused." Finally, Zippy's Friends (see Mishara & Ystgaard, 2006) can be considered "students' social-emotional development focused." We therefore concluded that all programs have at least one student-focused component in their intervention, but only two contain teacher-focused components (e.g., improving teachers' use of classroom rules and procedures). Remarkably, although the importance of establishing positive teacher–student relationships (our second classification) is emphasized in all programs, in none of the programs is this component explicitly integrated in the intervention or at least not in the descriptions of the interventions.

The duration of the intervention was categorized into three groups: less than 13 weeks, between 13 weeks and 1 year, and longer than 1 year. Dichotomous variables were added to indicate whether the study was conducted in the United States or in a different country (studies conducted in the United States were largely overrepresented). We included a variable indicating whether participants—students, classes/teachers, or schools—were randomly assigned to intervention and control groups.

The outcome measures were recoded into academic outcomes, behavior, social-emotional outcomes, motivation, and other outcomes. Scores on (standardized) tests, GPA, school grades, proficiency measures, academic competencies, and estimates of academic outcomes by the teacher were coded as academic outcomes. Concentration, attention, hyperactivity, externalizing problem behavior, internalizing problem behavior, aggression, conduct problems, antisocial behavior, obedience, problem solving behavior, self-control, and inhibition were coded

as behavioral outcomes. Social development, social skills, social competencies, emotional development, emotional skills, emotional competencies, emotion recognition, moral sensitivity, coping, emotion regulation, and empathy were coded as social-emotional outcomes. Academic motivation, school motivation, goal orientations, commitment to school, learning engagement, and enthusiasm were coded as motivational outcomes. All other, in our view, relevant student outcomes such as self-confidence, self-efficacy, peer acceptance, and time-on-task were coded as “other” outcomes.

Outcomes measured using highly unreliable instruments (Cronbach’s $\alpha < .40$) were not included. An additional categorical variable indicated whether the outcome measures were rated by the students (self-rating), by their teachers, by a researcher/observer, or by other people, usually parents or peers. We decided to include only those that were rated by the students themselves, the teachers, or the researchers/observers. This was done because we considered it to be more difficult for parents and peers to assess the students’ behavior in the classroom only, without taking behavior outside the classroom into account. The socioeconomic status of the students was recoded into more than 40% free or reduced lunch (low SES) or less than 40% free or reduced lunch (medium or high SES). The grade levels included in the studies were categorized into both lower and higher grades, pre–K to Grade 1, and Grade 2 and up. Finally, a dichotomous variable indicated whether regular students or the students with frequent problem behavior were assessed (i.e., despite the fact that the intervention was focused on the entire class).

Originally, we were interested in the results of the programs in follow-up tests. Follow-up tests were used in only 2 out of the selected 47 studies, and hence this variable could not be taken into account during the analyses. We were also interested in the educational context (e.g., whether it concerned instruction, independent seatwork, cooperative learning, or lesson transitions). Yet in most cases the intervention was implemented throughout the day rather than in specific educational contexts, or the educational context was not reported. Hence, this study characteristic could not be taken into account either. Furthermore, we aimed to include the classroom setting (group settings or frontal placement), but only four studies reported this information.

Data Analysis

Meta-analyses were mainly performed using the program Comprehensive Meta-Analysis of Biostat (CMA; Borenstein, Hedges, Higgins, & Rothstein, 2009). Only for the metaregression analyses with multiple predictors, we used the statistical program Hierarchical Linear Modeling Version 6, developed by Raudenbush, Bryk, and Congdon (2004). In a meta-analysis, the unit of analysis is not the individual participant but the effect size determined on the basis of primary studies’ outcomes. Therefore, an important part of the analyses is (re)calculation of the effect sizes, to enable a useful comparison between the reported effects of the different studies. In most of the intervention studies, the results data were based on a pretest–posttest control group design. Using the above-mentioned program, Hedges’s g was calculated. This is the adjusted standardized mean difference (d) between two groups, based on the pooled standard deviations. Hedges’s g is particularly useful for a meta-analysis of studies with different sample sizes. We

defined the direction of the effect in such a way that a positive effect size indicates that the intervention group did better than the control group (e.g., higher academic performance, better behavior) and a negative effect indicates that the control group did better than the intervention group. We defined the effects at the level of the students and not at the level of the class or school. Most of the data in the primary studies were also reported at the level of the students, but in 4.6% of the reported data the class or school was the unit of analysis. In these cases, we recomputed the class/school-level effect sizes by multiplying them by the square root of the intraclass correlation, as Hedges (2007) prescribes. Hedges and Hedberg (2007) reported an average intraclass correlation value (in models that corrected for pre-test scores) of about .1 for primary school students' performance in reading and mathematics. Average intraclass correlations for nonacademic outcomes were not reported. Our metastudy, however, included two studies (Benner, Nelson, Sanders, & Ralston, 2012; Raver et al., 2009) in which these were reported for behavioral outcomes, and both came to an average value of .1. This is also the value used in the meta-analyses of What Works Clearinghouse (2014). Therefore, we used .1 as the intraclass correlation value for all our recomputations.

For several interventions, multiple outcome measures of the same type were available. In these cases, we computed the intervention effect as the average effect of the multiple measures. CMA was also used to compute the variances of the individual interventions' effect sizes. This information was used to perform weighted analyses (with a random-effects model). The weight assigned to each intervention is the inverse of the variance. In this way, interventions with lower variances (which were the interventions with larger sample sizes) had a greater effect on the calculated summary effects.

The summary effects were estimated using a random-effects model. Moderator analyses (with analysis of variance [ANOVA] for meta-analytical data) were conducted using a mixed-effects model: The within subgroup effects were estimated with a random-effects model and the differences across subgroups with a fixed-effects model. In the analysis, the coded characteristics of CMS/CMP were modeled as predictors of the differences between the effects found. The predictors were categorized at the level of the intervention, and the dependent variables were the sizes of the effects (for all student outcomes) of these interventions. We wanted to perform a robust exploratory analysis of factors influencing the intervention effects. Therefore, we have examined many possible moderators, such as aspects related to the type of intervention, the duration of the intervention, the characteristics of the participants, and by whom the intervention was evaluated. With Hierarchical Linear Modeling, we examined the effects of the moderators simultaneously, by performing a metaregression analysis with a random-effects model. We included each measure of the interventions. So, if an intervention estimated the effect with three tests, all three measures were included. This enabled us to include the moderator "rater," which varied along each measure instead of along each intervention. We used the effects and variances that CMA calculated of the individual measures. However, to prevent interventions with multiple measures to be "overweighed," we adjusted the variances. We did this by multiplying the variance of each measure by the number of measures that were available for each intervention. So, if one intervention effect was measured using three tests, the

variances of each of the three measures were multiplied by three. Each measure as a result weighs only one third (the weight is $1/\text{variance}$) of its original weight. Combined, however, these three measures have the same weight as in CMA.

In this meta-analysis, we have performed multiple statistical significance tests. This raises the chance to conclude incorrectly that there is an effect (a Type I error): a false discovery. We corrected for this phenomenon by applying a Type I error correction method recommended by Polanin (2013) in his dissertation on addressing the issue of multiple testing in meta-analysis. Polanin (2013) advocates the “false discovery rate” (FDR) procedure described by Benjamini and Hochberg (1995) in combination with a “timeline of statistical significance testing” (see Polanin, 2013, p. 96). Applying this method balances the chances of incorrectly rejecting a hypothesis and incorrectly accepting a hypothesis (a Type II error). The FDR procedure in combination with the “timeline” is, in short, as follows: Let there be m tests for overall average effects and m between-groups tests. Using the “time line” means applying the FDR procedure within each group of tests separately. With m tests, we have m null hypotheses ($H_1, H_2, H_i, \dots, H_m$) and m p -values ($p_1, p_2, p_i, \dots, p_m$). Order the p values from low to high, start with the largest value, and find the largest p -value for which

$$p_i \leq \frac{i}{m} * \alpha,$$

where i is the ordered p value and α is the chosen level of control. We chose $\alpha = .05$. Then, reject the null hypotheses of p_i and smaller.

An elegant feature of CMA is that it is possible to examine the probability of biased results due to a phenomenon called *publication bias*. Studies are more likely to be published when the effects found in the study are significant or the study is based on a large sample size. Studies based on smaller sample sizes and reporting no significant effects might, therefore, be underrepresented in the meta-analysis. CMA is used to analyze the relationship between sample size and effect size. The program assumes that if there is a relationship between the two constructs, this can be attributed to missing studies. Furthermore, it estimates to what extent the results of the meta-analysis are likely to be biased.

Results

Characteristics of the Intervention Studies

We first present the descriptive characteristics of the selected studies. The results of 46 studies³ were used in the analyses, which together report the findings of 54 intervention studies. Table 1 gives an overview of the characteristics of the intervention studies.

The focus of most of the intervention studies was on changing the students’ (students’ behavior and/or students’ social-emotional development) and/or the teachers’ (i.e., their CMS) behavior through long-term interventions; the shortest intervention lasted 6 weeks and the longest 3 years. Only two intervention studies were explicitly focused on changing teacher–student relationships. A large variety of interventions was implemented in the studies. Only the PATHS program was implemented relatively often, in 10 intervention studies.

TABLE 1*Overview of the characteristics of the 54 intervention studies*

Intervention Characteristics	No. of interventions	% of interventions
Duration of the intervention		
<13 weeks	6	11.1
13 weeks to 1 year	30	55.6
>1 year	18	33.3
Focus of the intervention (an intervention can have more than one focus)		
Teachers' behavior	29	53.7
Students' behavior	46	85.2
Students' social-emotional development	40	74.1
Teacher–student relationship	2	3.7
Name of the intervention		
School-Wide Positive Behavior Support	3	5.6
Promoting Alternative Thinking Strategies	10	18.5
Good Behavior Game	4	7.4
Second Step	3	5.6
Zippy's Friends	3	5.6
Other	31	57.4
Country		
United States	39	72.2
Other	15	27.8
Grade years		
Pre–K and Grade 1	22	40.7
Grades 2–6	20	37.0
Both	12	22.2
Type of student sample		
Regular students	46	85.2
Students with behavior problems	5	9.3
Missing	3	5.6
Sex		
Girls	3	5.6
Boys	4	7.4
No results specification for students' sex	50	92.6
Socioeconomic status		
Low socioeconomic status	27	50.0
Mid and high socioeconomic status	15	27.8
Missing	12	22.2
Outcome variables (an intervention can have more than one outcome type)		
Academic outcomes	17 (37 tests)	31.5

(continued)

TABLE 1 (continued)

Intervention Characteristics	No. of interventions	% of interventions
Behavior outcomes	43 (147 tests)	79.6
Social-emotional outcomes	27 (58 tests)	50.0
Motivational outcomes	6 (10 tests)	11.1
Other outcomes	5 (10 tests)	9.3
Rater (total tests = 262)		
Teacher	137	52.3
Student	89	34.0
Observer	36	13.7

About three quarters of the intervention studies were conducted in the United States; the other studies were mainly conducted in European countries (Norway, Luxembourg, Belgium, the Netherlands, Germany, Denmark, Lithuania, Turkey, United Kingdom) and in Canada. Regarding the student sample characteristics, we found that both lower and higher grade levels were represented in the selected intervention studies and that regular students (without serious behavior problems) were commonly included. Although the socioeconomic status of the students was not indicated in several studies, we found that low-SES students were overrepresented in the selected studies compared with mid- and high-SES students. Three intervention studies reported results for boys and girls separately, and one intervention targeted boys only. The other 50 interventions did not distinguish their results according to students' sex.

Results were often reported for more than one outcome type. Table 1 shows how often each outcome was reported in total in our sample of interventions. Student behavior was by far the most common student outcome (44%), followed by social-emotional outcomes (28%) and academic outcomes (17%). In a few studies, student motivation (6%) or another outcome measure at the student level (5%; e.g., time-on-task, self-efficacy, peer acceptance) was reported. Also, intervention effects were often estimated using more than one measurement instrument. The total number of tests used in the interventions was 262. In half of these, the teachers rated the student outcomes, and in one third of the tests, student self-reports were used. In a few cases, an external observer rated the student outcomes.

Effects of the Interventions

The findings of meta-analytical analysis show that the classroom management interventions had a small but significant effect on various student outcome measures. Table 2 reports the statistics for all outcomes together, and for each outcome separately. These statistics are indices of the average effect sizes (Hedges's g), their variation (SE), and the source of variation: true differences or random error (I^2). The Q -statistics for the outcomes show if there is significant heterogeneity among the effect sizes. If so, it is likely that the interventions do not share the same true effect size. For the overall outcome, the Q -statistic indicates that this was the case, suggesting that the variations in effect size reflected real differences between the

TABLE 2
Effects of classroom management interventions

Outcome	Hedges's <i>g</i> (<i>SE</i>)	<i>Q</i> (degrees of freedom, <i>p</i>)	<i>I</i> ²	<i>T</i> ²
Overall	0.22 (0.02)**	342.45 (53, .00)**	84.52	0.01
Academic	0.17 (0.04)**	64.71 (16, .00)**	75.28	0.01
Behavior	0.24 (0.03)**	183.55 (42, .00)**	77.12	0.02
Social-emotional	0.21 (0.03)**	117.23 (26, .00)**	77.82	0.02
Motivation	0.08 (0.08)	16.00 (5, .01)*	68.74	0.02
Other	0.26 (0.10)*	11.08 (4, .03)*	63.90	0.03

p* < .05. *p* < .01.

interventions. *I*² indicates the percentage of the heterogeneity in intervention effect sizes that can be explained by differences between the interventions. Table 2 shows that *I*² for the overall effect was 84.52, which suggests that 84.52% of the dispersion of the interventions' effect sizes reflected real differences in effect size, and that 15.48% was due to random error. This also applied to each of the outcomes separately. *T*² is the estimated population variance of the effect sizes.

In an additional analysis, we examined whether the effect sizes differed significantly between the various groups of outcomes. This was found not to be the case (*Q*-between = 5.29, degrees of freedom = 4, *p* = .26).

The findings furthermore revealed that the meta-analysis was subject to some publication bias. Duval and Tweedie's Trim and Fill method (Borenstein et al., 2009; J. L. Peters, Sutton, Jones, Abrams, & Rushton, 2008) for a random-effects model showed that for all outcomes together, the meta-analysis lacked 12 interventions on the left side of the mean; this is a lower effect size than average. If these 12 interventions had been added, the average effect size would have been slightly lower with *g* = 0.17 (*SE* = 0.02). We also found publication bias for each outcome separately, except for the motivational outcomes. Duval and Tweedie's method indicated that for the academic, social-emotional, and "other outcomes" (e.g., time-on-task, self-efficacy, peer acceptance), interventions with lower effect sizes were lacking. For the behavioral outcomes, one intervention with a higher effect size was lacking. Figure 1 shows the funnel plots of the relationship between standard error and effect size for all outcomes together and for each outcome separately. The figures display the observed and imputed interventions. The imputed interventions are those that were estimated as probably lacking due to publication bias. The interventions with a small sample size generally have a larger standard error and appear at the bottom of the figure.

Moderator Analyses

We examined the relationship between the intervention effects and the type of classroom management intervention. Table 3 reports the average effects for each

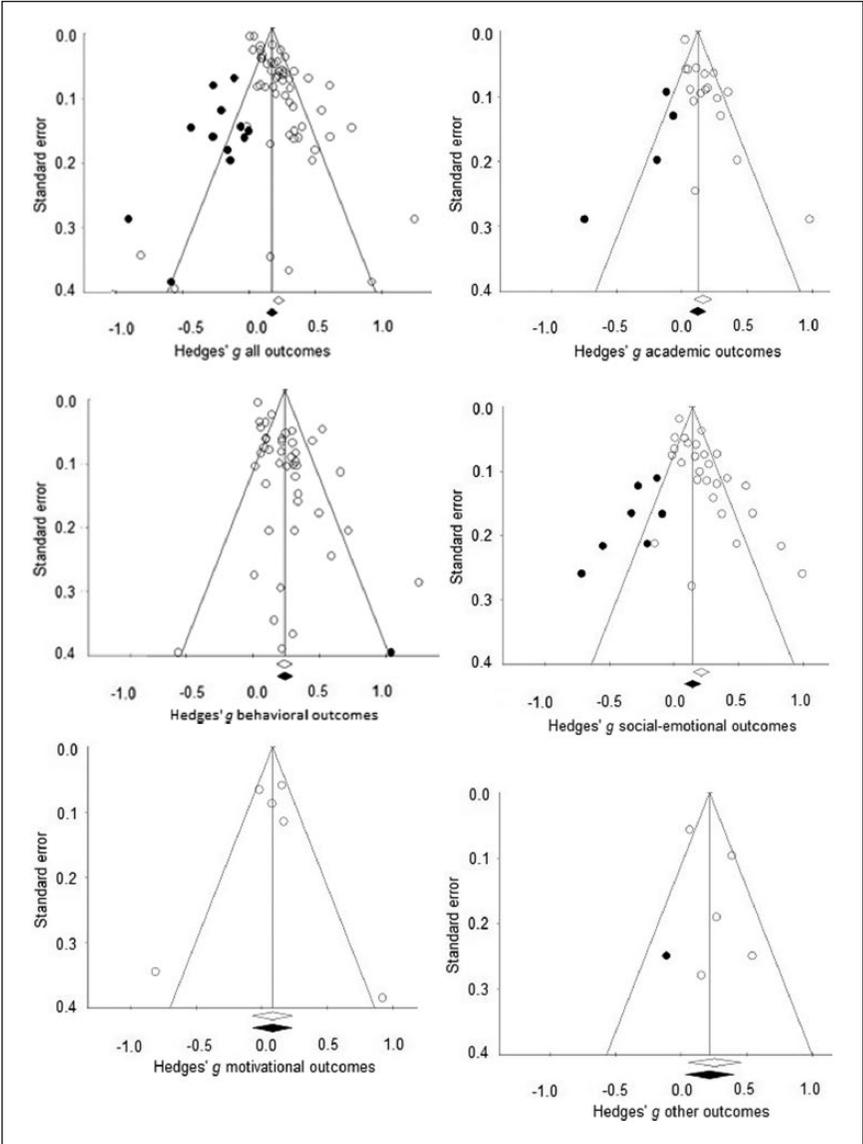


FIGURE 1. Funnel plots of standard error by effect size for the interventions. The observed interventions are represented by an open circle; imputed interventions are represented by a filled circle. The diamonds at the bottom represent the summary effect and its confidence interval, the open diamond for the observed interventions only, and the filled diamond for the observed and imputed interventions.

TABLE 3

Average effects (Hedges's g [SE]) for each focus component on the various outcome types

Focus component	Component included	Component not included	Q -between ^a (degrees of freedom, p)
All outcomes			
Teacher's behavior	0.20 (0.03)**	0.24 (0.03)**	0.88 (1, .35)
Students' behavior	0.21 (0.03)**	0.26 (0.05)**	0.85 (1, .36)
Students' social-emotional development	0.24 (0.02)**	0.15 (0.04)**	3.83 (1, .05)†
Teacher–student relationship	0.13 (0.09) ^b	0.22 (0.02)**	1.05 (1, .31)
Academic outcomes			
Teacher's behavior	0.21 (0.05)**	0.09 (0.03)** ^c	3.84 (1, .05)†
Students' behavior	0.18 (0.04)**	0.11 (0.06) ^c	0.86 (1, .35)
Students' social-emotional development	0.17 (0.03)**	0.15 (0.08)*	0.06 (1, .82)
Teacher–student relationship	0.24 (0.09)** ^b	0.16 (0.04)**	0.84 (1, .36)
Behavioral outcomes			
Teacher's behavior	0.21 (0.04)**	0.28 (0.04)**	1.46 (1, .23)
Students' behavior	0.23 (0.03)**	0.28 (0.10)**	0.24 (1, .63)
Students' social-emotional development	0.25 (0.03)**	0.20 (0.06)**	0.71 (1, .40)
Teacher–student relationship	0.06 (0.10) ^b	0.24 (0.03)**	2.92 (1, .09)
Social-emotional outcomes			
Teacher's behavior	0.16 (0.05)**	0.24 (0.04)**	1.92 (1, .17)
Students' behavior	0.20 (0.04)**	0.25 (0.05)**	0.64 (1, .42)
Students' social-emotional development	0.25 (0.03)**	0.04 (0.02)* ^c	30.35 (1, .00)**
Teacher–student relationship	0.06 (0.09) ^b	0.22 (0.03)**	2.99 (1, .08)
Motivational outcomes			
Teacher's behavior	0.08 (0.09) ^b	0.08 (0.11)	0.00 (1, .98)
Students' behavior	0.08 (0.08)	—	—
Students' social-emotional development	0.14 (0.05)** ^c	0.01 (0.37) ^c	0.12 (1, .73)
Teacher–student relationship	0.08 (0.09) ^b	0.08 (0.11)	0.00 (1, .98)
Other outcomes			
Teacher's behavior	0.38 (0.08)** ^c	0.07 (0.06) ^b	10.23 (1, .00)**
Students' behavior	0.39 (0.09)** ^c	0.09 (0.06) ^b	7.67 (1, .01)**
Students' social-emotional development	0.18 (0.10) ^c	0.39 (0.10)** ^b	2.39 (1, .12)
Teacher–student relationship	0.27 (0.19) ^b	0.26 (0.12)* ^c	0.00 (1, .95)

^aOne-way analysis of variance for meta-analysis. ^bStatistic in cell is based on only one or two interventions. ^cStatistic in cell is based on no more than three or four interventions.

† $p = .05$. * $p < .05$. ** $p < .01$.

TABLE 4*Average effects (Hedges's g [SE]) for number of focus components*

Outcome	1 component	2 components	3 or 4 components	Q -between ^a (degrees of freedom, p)
Overall	0.17 (0.07)**	0.24 (0.03)**	0.20 (0.04)**	1.19 (2, .55)
Academic	0.11 (0.04) ^b	0.10 (0.06)	0.23 (0.03)**	7.35 (2, .03) ^c
Behavior	0.27 (0.10)**	0.24 (0.04)**	0.20 (0.04)**	0.76 (2, .68)
Social-emotional	0.16 (0.06)**	0.27 (0.05)**	0.16 (0.08)*	3.06 (2, .22)
Motivation	0.01 (0.37) ^b	0.16 (0.05)** ^d	0.08 (0.09) ^d	0.72 (2, .70)
Other	0.07 (0.06) ^d	0.36 (0.09)** ^d	0.37 (0.15)* ^d	9.70 (2, .01)** ^c

a. One-way analysis of variance for meta-analysis. ^bStatistic in cell is based on no more than three or four interventions. ^cEffect is likely a "false discovery." ^dStatistic in cell is based on only one or two interventions.

* $p < .05$. ** $p < .01$.

component of the interventions that we distinguished based on their focus. The table presents the estimated effects for interventions that include a particular component ("component included") and for interventions that do not include a particular component ("component not included"). As interventions can focus on multiple components at once, we also examined whether the effectiveness of the intervention depended on the number of components it addressed. Table 4 reports these results. In addition, Table 5 shows the effects for all the combinations of components that were present in our meta-analysis, to indicate whether a particular combination of certain components was more effective. Last, Table 6 shows the effects for five specific intervention programs of which our meta-analysis included at least three studies. A sixth category contained the other interventions (i.e., all other interventions in our meta-analysis, thus those that did not focus on SWPBS, PATHS, GBG, Second Step, or Zippy's Friends). Using a one-way ANOVA model for meta-analyses, we tested for each outcome separately whether the differences in effects were significant. The Q -betweens (which follow the same logic as an F -value in regular ANOVA) are reported in the last columns of the tables.

Table 3 shows that for all outcome types together, interventions were not more effective when they focused on changing the teachers' behavior (e.g., keeping order, introducing rules and procedures), changing student behavior (either students' behavior or students' social-emotional development, or both), or improving the teacher-student relationship. However, with a p value of exactly .05, the results do suggest a trend that focusing on the social-emotional development of students had an effect. Programs that addressed this component had a slightly higher effect size than programs that did not. Taking a closer look at the different types of outcomes, it can be seen that particularly the social-emotional outcomes (e.g., empathy for other children's feelings) benefitted from programs designed to enhance students' social-emotional development. Furthermore, we found a trend that academic outcomes seemed to benefit from a program focused on improving

TABLE 5
Average effects (SE) for focus of classroom management interventions

Focus	Overall	Academic	Behavior	Social-emotional	Motivation	Other
Student behavior ($N = 3$)	0.06 (0.37) ^a	0.18 (0.07)** ^b	0.21 (0.07)** ^b	0.01 (0.07) ^b	0.01 (0.37) ^a	—
Student social-emotional (including Zippy's Friends; $N = 4$)	0.20 (0.05)** ^a	0.08 (0.04) ^b	0.28 (0.13)* ^a	0.19 (0.06)** ^a	—	0.07 (0.06) ^b
Student behavior + student social-emotional (including PATHS and Second Step; $N = 18$)	0.27 (0.04)**	0.05 (0.06) ^b	0.27 (0.05)**	0.29 (0.04)**	0.16 (0.05)** ^b	0.16 (0.28) ^b
Teacher + student behavior (including SWPBS and GBG; $N = 11$)	0.16 (0.04)**	0.16 (0.10) ^a	0.19 (0.07)**	0.04 (0.02)* ^b	—	0.39 (0.10)** ^b
Teacher + student social-emotional ($N = 3$)	0.37 (0.09)** ^a	—	0.29 (0.10)** ^b	0.36 (0.10)** ^a	—	—
Teacher + student behavior + student social-emotional ($N = 13$)	0.20 (0.04)**	0.23 (0.04)**	0.22 (0.05)**	0.18 (0.10)	—	0.54 (0.25)* ^b
Teacher + relation + social-emotional ($N = 1$)	0.31 (0.20) ^b	0.43 (0.20)* ^b	0.21 (0.21) ^b	—	—	0.27 (0.19) ^b
All components ($N = 1$)	0.09 (0.10) ^b	0.20 (0.09)* ^b	0.02 (0.12) ^b	0.06 (0.09) ^b	0.08 (0.09) ^b	—
Q -between ^c (degrees of freedom, p)	8.76 (7, .27)	13.34 (6, .04)* ^d	5.09 (7, .65)	42.98 (6, .00)**	0.72 (2, .70)	11.08 (4, .03)* ^d
Q -between ^c (degrees of freedom, p) restricted	6.85 (5, .23)	0.44 (1, .51)	1.27 (3, .74)	3.55 (3, .32)	—	—

Note. PATHS = Promoting Alternative Thinking Strategies, GBG = Good Behavior Game, SWPBS = School-Wide Positive Behavior Support.

^aStatistic in cell is based on no more than three or four interventions. ^bStatistic in cell is based on only one or two interventions. ^cOne-way analysis of variance for meta-analysis. Restricted analysis of variance is based on the cells with three or more interventions. ^dEffect is likely a "false discovery."
* $p < .05$. ** $p < .01$.

TABLE 6
Average effects (SE) for specific classroom management interventions

Outcome	SWPBS	PATHS	GBG	Second Step	Zippy's Friends	Other	Q -between ^a (degrees of freedom, p)
Overall	0.03 (0.02) ^b	0.29 (0.05)**	0.22 (0.09) ^{*b}	0.21 (0.05) ^{**b}	0.19 (0.08) ^{*b}	0.23 (0.03)**	46.32 (5, .00)**
Academic	0.01 (0.01) ^c	—	0.09 (0.11) ^c	—	0.11 (0.06) ^{*c}	0.19 (0.04)**	24.08 (3, .00)**
Behavior	0.16 (0.13) ^c	0.26 (0.06)**	0.25 (0.09) ^{**b}	0.22 (0.05) ^{**b}	0.18 (0.08) ^{*b}	0.25 (0.05)**	1.24 (5, .94)
Social-emotional	0.04 (0.02) ^{*c}	0.32 (0.05)**	—	0.16 (0.07) ^{*c}	0.22 (0.09) ^{*b}	0.20 (0.05)**	34.28 (4, .00)**
Motivation	—	0.17 (0.11) ^c	—	—	—	0.06 (0.10)	0.52 (1, .47)
Other	—	—	—	0.16 (0.28) ^c	0.07 (0.06) ^c	0.38 (0.08) ^{**b}	10.32 (2, .01)**

Note. PATHS = Promoting Alternative Thinking Strategies, GBG = Good Behavior Game, SWPBS = School-Wide Positive Behavior Support. ^aOne-way analysis of variance for meta-analysis. ^bStatistic in cell is based on no more than three or four interventions. ^cStatistic in cell is based on only one or two interventions.

* $p < .05$. ** $p < .01$.

teachers' classroom management and their behavior; here, the p value was again exactly .05. The category "other outcomes" showed positive effects for teacher-focused and students' behavior-focused programs. Yet these results were based on very few interventions and should, therefore, be interpreted with care.

Table 4 indicates that academic outcomes were higher when interventions were focused on three or all components. The category "other outcomes" showed higher effects for interventions with at least two components. The number of components had no effect on the remaining outcome types. The effects on the academic outcomes and on the category "other outcomes" are, however, likely a "false discovery." In this meta-analysis, we have performed many moderator analyses. This raises the chance to conclude that there is an effect where in fact there is no effect (a Type I error). Therefore, we lowered the maximum p -values for significance (for more details see the section "Data Analysis").

As shown in Table 5, we analyzed the differences between the various combinations of focus components in two ways: based on all categories and based on the categories with three or more interventions (the restricted ANOVA). The latter analysis has the advantage that the number of groups in the analysis is more in line with the number of interventions included. According to Borenstein et al. (2009), a meta-analysis should include no more than one group per approximately 10 interventions. As such, the results of the restricted ANOVA results are to be preferred to the results of the analysis based on all categories. The results of the restricted ANOVA suggest that none of the combinations of focus components of interventions makes a difference. An interesting descriptive finding was that the most common combinations of classroom management components were programs combining a focus on students' behavior and students' social-emotional development (18 studies) and programs combining these two student components with a teacher focus (13 studies). Slightly less common were programs that combined a focus on students' behavior and a focus on teachers (11 studies). Other combinations of components were less frequently observed (5 different combinations across 12 studies).

Table 6 reveals that there were differences in effectiveness between the specific programs, except for the behavioral and motivational outcomes. When we focused on all outcome types together, we found all programs to have small to moderate effects. Only SWPBS had no effect. The specific programs seemed less effective than the category "other interventions" for academic and "other" outcomes. PATHS was found to have the highest effect on social-emotional outcomes, and SWPBS the lowest. Again, the results should be interpreted with care, as some averages are based on very few intervention studies.

The next moderator analyses were focused on differences related to student characteristics. Table 7 reports the statistics for sex, grade year, socioeconomic status, student behavior, and country. None of the reported student characteristics were found to cause differences in the intervention effects. We found a difference for socioeconomic status and for country only on "other outcomes," but these analyses were based on a very small number of interventions and should be interpreted with care. We also investigated whether the intervention effect was related to the duration of the program (see Table 7). Again, we found hardly any differences between the moderator variable and the intervention effects. We found a

TABLE 7
Average effects (SE) for sex, grade year, socioeconomic status, student behavior, country, duration, and rater

Intervention Characteristics	Overall	Academic	Behavior	Social-emotional	Motivation	Other
Sex						
Girls	0.10 (0.09) ^a	0.06 (0.09) ^b	0.30 (0.09)** ^a	-0.01 (0.10) ^b	0.02 (0.10) ^b	0.23 (0.28) ^b
Boys	0.23 (0.11)** ^a	0.54 (0.32) ^b	0.19 (0.09) ^a	0.02 (0.09) ^b	-0.05 (0.09) ^b	0.31 (0.26) ^b
<i>Q</i> -between ^c (degrees of freedom, <i>p</i>)	0.88 (1, .35)	2.17 (1, .14)	0.87 (1, .35)	0.03 (1, .86)	0.24 (1, .63)	0.05 (1, .82)
Grade year						
Pre-K and Grade 1	0.28 (0.04)**	0.23 (0.04)**	0.27 (0.05)**	0.25 (0.06)**	0.11 (0.07) ^b	0.39 (0.10)** ^b
Grades 2-6	0.17 (0.03)**	0.15 (0.06)**	0.20 (0.05)**	0.21 (0.05)**	0.05 (0.07) ^b	0.24 (0.23) ^b
Both	0.20 (0.04)**	0.09 (0.06) ^a	0.25 (0.06)**	0.19 (0.07)**	0.07 (0.09) ^b	0.24 (0.16) ^b
<i>Q</i> -between ^c (degrees of freedom, <i>p</i>)	4.33 (2, .12)	4.45 (2, .11)	1.16 (2, .56)	0.45 (2, .80)	0.14 (2, .93)	0.87 (2, .65)
Socioeconomic status						
Low SES ^d	0.20 (0.03)**	0.15 (0.05)**	0.21 (0.04)**	0.21 (0.04)**	0.13 (0.05)** ^b	0.38 (0.08)** ^a
Mid and high SES ^d	0.21 (0.04)**	0.18 (0.04)** ^a	0.24 (0.04)**	0.14 (0.07)*	-0.02 (0.07) ^b	0.07 (0.06) ^b
<i>Q</i> -between ^{c,d} (degrees of freedom, <i>p</i>)	0.02 (1, .88)	0.34 (1, .56)	0.23 (1, .63)	0.95 (1, .33)	3.35 (1, .07)	10.23 (1, .00)**
Student behavior						
Regular	0.20 (0.02)**	0.16 (0.04)**	0.22 (0.02)**	0.19 (0.03)**	0.08 (0.08)	0.26 (0.10)*
Behavior problems	0.27 (0.08)**	0.50 (0.44) ^b	0.29 (0.08)**	—	—	—
<i>Q</i> -between ^{c,d} (degrees of freedom, <i>p</i>)	0.62 (1, .43)	0.59 (1, .44)	0.72 (1, .40)	—	—	—

(continued)

TABLE 7 (continued)

Intervention Characteristics	Overall	Academic	Behavior	Social-emotional	Motivation	Other
Country						
United States	0.20 (0.03)**	0.16 (0.04)**	0.21 (0.03)**	0.20 (0.04)**	0.08 (0.08)	0.38 (0.08)** ^a
Other country	0.26 (0.05)**	0.18 (0.07)** ^b	0.28 (0.05)**	0.25 (0.07)**	—	0.07 (0.06) ^b
<i>Q</i> -between ^c (degrees of freedom, <i>p</i>)	1.31 (1, .25)	0.03 (1, .85)	1.24 (1, .27)	0.40 (1, .53)	—	10.23 (1, .00)**
Duration						
<13 weeks	0.19 (0.24)	0.30 (0.13)** ^b	0.29 (0.21) ^a	0.83 (0.22)** ^b	0.05 (0.87) ^b	—
13 weeks–1 year	0.23 (0.03)**	0.16 (0.05)**	0.21 (0.03)**	0.21 (0.04)**	0.11 (0.07) ^b	0.23 (0.12) ^a
>1 year	0.21 (0.03)**	0.16 (0.06)**	0.25 (0.05)**	0.19 (0.06)**	0.07 (0.09) ^b	0.37 (0.19) ^b
<i>Q</i> -between ^c (degrees of freedom, <i>p</i>)	0.18 (2, .92)	1.10 (2, .58)	0.56 (2, .76)	8.44 (2, .02)** ^e	0.14 (2, .93)	0.38 (1, .54)
Rater						
Teacher	0.24 (0.03)**	0.12 (0.03)**	0.26 (0.03)**	0.24 (0.04)**	0.09 (0.04)** ^a	0.16 (0.06)** ^b
Student	0.16 (0.03)**	0.16 (0.04)**	0.10 (0.03)**	0.18 (0.05)**	0.10 (0.38) ^a	0.18 (0.16) ^a
Observer	0.30 (0.07)**	—	0.26 (0.07)**	0.47 (0.49) ^b	—	0.39 (0.10)**
<i>Q</i> -between ^c (degrees of freedom, <i>p</i>)	6.27 (2, .04)** ^e	0.37 (1, .54)	13.44 (2, .00)**	1.06 (2, .60)	0.00 (1, .99)	4.38 (2, .11)

^aStatistic in cell is based on no more than three or four interventions. ^bStatistic in cell is based on only one or two interventions. ^cOne-way analysis of variance for meta-analysis.

^dInterventions with missing data were excluded from the analysis. ^eEffect is likely a “false discovery.”

p* < .05. *p* < .01.

TABLE 8

Average effects (B [SE]) for focus components, and moderators, duration, grade year, socioeconomic status, student behavior, country, and rater

Intervention Characteristics	Overall	Academic	Behavior	Social-emotional
Intercept	.11 (.09)	.13 (.20)	.18 (.11) [†]	-.02 (.21)
Focus teacher behavior	.03 (.04)	.27 (.13)*	.04 (.07)	-.00 (.11)
Focus student behavior	-.01 (.04)	-.13 (.15)	-.02 (.07)	-.01 (.07)
Focus student social-emotional	.11 (.04)*	.09 (.09)	.15 (.08) [†]	.30 (.15) [†]
Focus relation teacher–student	-.12 (.10)	-.13 (.17)	-.13 (.11)	-.19 (.16)
<13 Weeks	.03 (.05)	-.14 (.19)	-.05 (.15)	.51 (.26) [†]
13 Weeks to 1 year	-.06 (.04)	-.16 (.09) [†]	-.15 (.08) [†]	-.07 (.10)
Grade maximum 1	.07 (.06)	.09 (.08)	.07 (.08)	-.01 (.13)
Grades 2–6	-.04 (.04)	-.12 (.13)	-.04 (.07)	-.03 (.10)
SES low	-.04 (.04)	-.08 (.10)	-.06 (.08)	-.02 (.13)
Behavior problems	.11 (.09)	.11 (.13)	.13 (.12)	—
Not United States	.06 (.05)	-.01 (.14)	.04 (.07)	-.01 (.11)
Rater teacher	.03 (.04)	.15 (.09)	.01 (.05)	.08 (.14)
Rater student	-.03 (.05)	—	-.05 (.07)	.01 (.16)

Note. SES = socioeconomic status. Reference category for focus components is “not included,” for duration “more than 1 year,” for grade “all grades,” for SES “mid and high SES + SES missing,” for student behavior “regular + missing,” for country “United States,” and for rater “observer.”

[†] $p < .10$. * $p < .05$. ** $p < .01$.

small difference for the social-emotional outcomes, but this effect was likely to be a “false discovery.” The estimated intervention effect might relate to how the effect was measured. In many intervention studies, the effect was estimated using ratings by the teacher, the student, or an observer (see Table 7). We found significant differences between the raters for all outcomes together and for behavioral outcomes. Students reported less improvement after following the program in comparison with reports filled in by teachers and observers.

Finally, we examined the influence of the moderators on the intervention effect in a multiple metaregression model. We included the four focus components, the moderators duration, grade, SES, student behavior, country, and rater. We excluded sex, as only very few interventions distinguished between boys and girls. This analysis has the advantage that the effects of the moderators are analyzed simultaneously and that it shows the unique contribution of each moderator, while taking the other moderator effects into account. Although our metastudy did not include a sufficient number of interventions to maintain high power of the analysis when including the multiple moderators, we believe the results are informative. Table 8 presents the metaregression models for the various outcome types. We were, however, unable to run the models for the outcome categories “motivation” and “other” because of the very low numbers of interventions. In general, we found the same results as in the previous analyses, but the models did not have a good fit. Therefore, we did not check which effects were likely true and which

were likely a false discovery (Polanin, 2013). The fact that the findings do support our results presented earlier does, however, strengthen our conclusions.

Discussion

Summary of the Results

The meta-analysis included 54 classroom management interventions (presented in 47 different studies) aimed at enhancing students' academic, behavioral, social-emotional, motivational, or other related student outcomes. A large variety of interventions was implemented in the studies that met our inclusion criteria. Our analyses included five classroom management interventions that were implemented in at least three studies, namely, SWPBS, PATHS, GBG, Second Step, and Zippy's Friends. However, together they represented only 43% of the overall sample of selected studies.

In 85% of the included studies, interventions were used that (among other foci) focused on changing students' behavior, and 74% at least partly focused on improving students' social-emotional development. Half of the included studies reported on interventions that (at least partly) focused on changing the teachers' behavior (54%). Only two intervention studies were explicitly focused on improving teacher–student relationships (4%). The most common classroom management components were a combination of focusing on students' behavior and students' social-emotional development (18 studies), and these two student components combined with a teacher focus (13 studies). This trend toward more student-centered approaches rather than teacher-centered approaches is in line with the general tendency in primary education toward student-centered learning environments.

Across all interventions, we calculated an overall effect of $g = 0.22$ on the various student outcomes (0.17 if the publication bias is taken into account). There were no significant differences between the various groups of outcomes: academic, behavioral, social-emotional, motivational, and other (e.g., time-on-task, self-efficacy, peer acceptance). Thus, the results of the meta-analysis confirm the finding of generally positive effects of classroom management interventions on student outcomes in primary education. In prior meta-analyses (Durlak et al., 2011; Marzano et al., 2003; Oliver et al., 2011), the reported effects were generally similar in size (i.e., when the effect sizes measured at the classroom level, e.g., Marzano et al., 2003, or measured at the school level, e.g., Oliver et al., 2011, are recalculated). Durlak et al. (2011) found somewhat larger effects for social-emotional outcomes (0.57) than we found in our study (0.22). Our meta-analysis included the recent literature only (published between 2003 and 2013). It is, therefore, noteworthy that our overall finding that classroom management interventions are generally effective in enhancing student outcomes is in line with the findings of prior meta-analyses, which were mostly based on earlier publications.

To determine to which components of the classroom management interventions their effectiveness can be attributed, we performed several moderator analyses. The results indicated that interventions focused on the social-emotional development of the students were somewhat more effective than interventions

without this component. In particular, the social-emotional outcomes benefitted from this component. This also applied to the outcome category “other” (e.g., time-on-task, self-efficacy). Furthermore, the exact combination of components on which programs focused had no influence on the intervention effect. We examined the effectiveness of the five intervention programs that were most common in our meta-analysis. We found that all programs were equally effective, except for SWPBS, which was not found to have an effect on the outcome measure “all outcome types together.”

Additional moderator analyses revealed no large differences in the reported effects with respect to sex, socioeconomic status (low vs. mid or high), student behavior (regular or students with behavioral problems), grade year (pre-K to 1, 2–6, or both), or country (United States vs. non-United States), indicating that all students may benefit from a classroom management intervention.

Scientific Contribution

The findings of the present meta-analysis contribute to the current body of knowledge on classroom management by bringing together a broad span of recently conducted intervention studies on classroom management. In the selected studies, appropriate research designs were used to investigate the effects of various CMS/CMP on a variety of student outcomes. Whereas most prior researchers included studies without control groups in their meta-analyses, our focus was solely on studies with a control group. Therefore, maturation effects on social-emotional development, behavior, and achievement were controlled for in designs with a control group. Hence, we can be confident that the reported effects on student outcomes were caused by the interventions. Moreover, a range of different student outcomes were used: academic, behavioral, social-emotional, motivational, and other relevant student outcomes. The fact that many studies included multiple outcome measures enabled us to evaluate the effects of the interventions on (almost) all these outcomes.

Another relevant point is that the studies we included were published in the past decade, and thus in current educational settings. In some studies, the data used were collected several years earlier; however, in most studies, the data were collected in relatively modern classrooms. Furthermore, we paid specific attention to CMP that are commonly used in educational practice (SWPBS, PATHS, GBG, Second Step, and Zippy’s Friends). As yet, the effectiveness of several of these programs has not been investigated intensively. Although only a small number of studies of these programs could be included in our analyses (a minimum of three studies per program), we found that all programs (except SWPBS for “all outcome types together”) positively enhanced student outcomes.

Practical Implications

Classroom management aims to facilitate both academic and social-emotional learning (Evertson & Weinstein, 2006). In our meta-analysis, the strongest effects were found for programs targeting social-emotional development, particularly on the social-emotional outcome measure. This is considered a promising finding given that in current society social skills are important for success later in the school career and in the work force (Jennings & DiPrete, 2010; S. A. Lynch &

Simpson, 2010; Rhoades, Warren, Domitrovich, & Greenberg, 2011). Jennings and DiPrete (2010), for example, found that social and behavioral skills have a positive effect on the growth of academic skills in the early elementary grades. We would like to stress that understanding the link between classroom management and social-emotional development seems to be of particular importance for (student) teachers. Better social and emotional skills have positive effects on various educational outcomes at the individual student level. Moreover, at the classroom and school levels, positive effects may be expected as well although this aspect was not part of our study. For example, the atmosphere in the classroom may improve when individual students are better able to work together in groups and are better at solving problems without interference of the teacher. When teachers decide to implement a particular classroom management intervention in their classrooms, the program should therefore at least focus on students' social-emotional development: This has proven to be effective on various student outcomes.

A second finding of this meta-analysis was that for the interventions that focused on changing teachers' classroom management (e.g., keeping order, introducing rules and procedures, disciplinary interventions), we found a tentative result that these interventions had a small effect on students' academic outcomes ($p = .05$). Classroom management is considered a precondition for learning; effective teaching, and learning cannot take place in poorly managed classrooms (V. F. Jones & Jones, 2012). These findings can be explained through improved time-on-task, improved instruction practices, and increased opportunity-to-learn, but this hypothesized causal chain needs to be further explored and validated in future research. Time-on-task was one of the outcomes we classified in the category "other outcome measures." Because this outcome measure was used in only a few studies, it was not feasible to analyze it separately. The category "other outcome measures" also included outcome measures such as self-efficacy and peer acceptance. More work is needed to understand how exactly student learning can be maximized through classroom management.

It must be remembered that most interventions (on average) showed positive effects on all student outcomes. Our findings clearly indicate that all students may benefit from these interventions. It is, however, essential that all stakeholders (policymakers, principals, teachers, and teacher educators) realize that the programs we investigated were often school-wide approaches in which a broad variety of strategies was used. This indicates that there is no simple solution for classroom management problems.

All in all, we would like to stress the importance of having a strong focus on classroom management in every primary school and classroom: Our study showed that all students may benefit from it. Implementation of effective classroom management interventions could be further stimulated (e.g., by the government) by providing schools with adequate information on those interventions with strong evidence on their effectiveness and those without. Moreover, teacher training programs should, in our view, integrate the existing knowledge about effective classroom management more strongly into their programs. By doing so, they can train their student teachers to manage classrooms effectively. Improving current teachers' classroom management skills is another element to incorporate.

As our results showed, it is very plausible that this will increase students' academic outcomes.

Limitations and Suggestions for Further Study

The studies included in the meta-analysis predominantly reported on the effectiveness of school-wide programs that had a broad focus on improving teaching practices, teacher–student relationships, student behavior, and student social-emotional development. Although the effects of school-wide universal CMP have often been investigated, few researchers have used pretest–posttest control group designs to estimate the effects on students' learning (both academic and social-emotional) and/or student behavior (see also Chitiyo, May, & Chitiyo, 2012). Consequently, the number of studies with a broad focus that met our inclusion criteria was small, considering that 241 potential studies resulted from the literature search. Although the number of studies included was sufficient for the analyses, we would like to stress that the results should be interpreted with some caution. The findings showed that our meta-analysis was subject to some publication bias, in particular for the categories “all outcomes” and for the “social-emotional outcomes.” A possible explanation for this bias relates to our search criteria. We focused on peer-reviewed journal articles and abstract collections. Studies on interventions with no significant effects are less likely to be published. In addition, studies described in theses and dissertations were excluded because these were not peer-reviewed. Although our search criterion has also the potential of neglecting some studies, we advocate that it is a useful criterion for a first selection of studies of sufficient quality.

Another limitation is that the findings of moderator analyses showed that students reported less enhancement on behavior by the interventions than was reported by teachers and observers, which might be caused by teachers' and observers' desire to find significant progress. Then again, self-reports of young students may be inaccurate if the research instruments are too complicated for them. Furthermore, we were unable to take all moderators into account in one single analysis and find a good model fit. This is probably due to the relatively low number of studies that met the inclusion criteria for the meta-analysis.

With regard to the outcome measures, we would like to stress that various measures were used, for instance, for academic outcomes. The use of standardized tests was limited, which makes it difficult to generalize the results to all academic outcomes. Time-on-task, which we expected to be a relevant outcome measure, was not often measured. Furthermore, various instruments were used to measure student behavior and students' social-emotional outcomes. Although we eliminated student outcomes measured using highly unreliable instruments, the construct validity of the various instruments was often unclear. As we have mentioned a number of times above, our results need to be interpreted with care.

A recommendation for further research pertains to the use of longitudinal studies. Out of the 241 potentially suitable publications, we found only two studies in which the long-term effects of a classroom management intervention (GBG) were measured: that is, the effects of implementing the intervention in Grades 1 and 2 on student outcomes during adolescence (Bradshaw, Zmuda, Kellam, & Lalongo, 2009; Kellam et al., 2008). More longitudinal studies are needed to investigate the

maintenance effects of classroom management interventions, for example, by using follow-up tests on various student outcomes at different ages. In particular, the school-wide universal CMP may have sustained effects on students' behavior and social-emotional development, because these are relatively intensive programs.

Finally, we would like to present some recommendations for the scientific community on the basis of our experiences in reporting pretest–posttest control group designs used to evaluate the effectiveness of classroom management interventions. We found that numerous studies lacked detailed descriptions of the intervention that was implemented in the schools (e.g., specific focus of the teacher sessions and/or student sessions, type of training teachers and/or students received, and duration of the intervention). Moreover, very few studies reported the classroom setting (e.g., group or frontal placement) in which the intervention was implemented, whereas such contextual factors may strongly influence student behavior in the classroom. Similarly, it was often unclear within what type of school or educational context (e.g., during instruction, collaborative assignments, independent seatwork, or throughout the school day) the intervention was implemented. And when the intervention was implemented throughout the school day, it was unclear how the school days were normally organized (e.g., the amount of instruction time, independent seatwork, how often students worked collaboratively in groups, whether some students followed an individual learning trajectory, whether computers were used throughout the day, and whether teaching assistants were present). Information on these aspects makes the interpretation of the effectiveness of classroom management interventions much more insightful and, moreover, makes the findings much easier to replicate. We therefore strongly recommend including detailed descriptions of these aspects in scientific papers evaluating the effectiveness of CMS/CMP. Another recommendation is to provide detailed information on the research design and sampling procedures. On several occasions, it was unclear (a) whether a control group was used, (b) how the randomization or matching across intervention and control groups was performed, and (c) whether the students were representative of the student population (e.g., many studies lacked details on gender, socioeconomic status, or ethnicity of the students included). In reporting the results, mean scores, standard deviations, and sample sizes among intervention and control groups should be reported for both pretest and posttest measures. Only then can effect sizes be properly calculated. Moreover, for these measures, reliable and validated research instruments should be used (and information about this should be reported).

Despite the aforementioned limitations and the clear need for more high-quality program evaluations, sufficient evidence was found that several classroom management interventions lead to different types of outcomes for these interventions to be considered for implementation in primary school classrooms. As a result of this meta-analysis, preconditions for effective teaching and learning found in recent studies have been identified.

Notes

This work is part of the research program, NWO Review Studies, which is partly financed by the Netherlands Organisation for Scientific Research (NWO).

¹The studies of Holen, Waaktaar, Lervåg, and Ystgaard (2012) and Holen, Waaktaar, Lervåg, and Ystgaard (2013) were counted as one study, because they reported the results

of the same intervention study. The only difference between the studies was the outcome variables reported.

²We referred to these programs as classroom management programs; however, we acknowledge that not all programs have presented themselves in such terms. The inclusion of these programs follows the broad definition of classroom management of Evertson and Weinstein (2006).

³The studies by Holen et al. (2012) and Holen et al. (2013) were counted as one study, because they reported the results of the same intervention study. The only difference between the studies was the outcome variables reported.

References

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

- *Arda, T. B., & Ocak, Ş. (2012). Social competence and Promoting Alternative Thinking Strategies—PATHS preschool curriculum. *Educational Sciences: Theory & Practice, 12*, 2691–2698.
- Barrish, H. H., Saunders, M., & Wolf, M. M. (1969). Good behavior game: Effects of individual contingencies for group consequences on disruptive behavior in a classroom. *Journal of Applied Behavior Analysis, 2*, 119–124. doi:10.1901/jaba.1969.2-119
- *Bavarian, N., Lewis, K. M., DuBois, D. L., Acock, A., Vuchinich, S., Silverthorn, N., . . . Flay, B. R. (2013). Using social-emotional and character development to improve academic outcomes: A matched-pair, cluster-randomized controlled trial in low-income, urban schools. *Journal of School Health, 83*, 771–779. doi:10.1111/josh.12093
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological), 57*, 289–300.
- *Benner, G. J., Nelson, J. R., Sanders, E. A., & Ralston, N. C. (2012). Behavior intervention for students with externalizing behavior problems: Primary-level standard protocol. *Exceptional Children, 78*, 181–198.
- *Bierman, K. L., Coie, J. D., Dodge, K. A., Greenberg, M. T., Lochman, J. E., McMahon, R. J., & Pinderhughes, E. (2010). The effects of a multiyear universal social-emotional learning program: The role of student and school characteristics. *Journal of Consulting and Clinical Psychology, 78*, 156–168. doi:10.1037/a0018607
- *Bierman, K. L., Domitrovich, C. E., Nix, R. L., Gest, S. D., Welsh, J. A., Greenberg, M. T., . . . Gill, S. (2008). Promoting academic and social-emotional school readiness: The Head Start REDI program. *Child Development, 79*, 1802–1817. doi:10.1111/j.1467-8624.2008.01227.x
- Bierman, K. L., Greenberg, M. T., & the Conduct Problems Prevention Research Group. (1996). Social skills training in the fast track program. In R. D. Peters & R. J. McMahon (Eds.), *Preventing childhood disorders, substance abuse, and delinquency* (pp. 65–89). Thousand Oaks, CA: Sage.
- Bohn, C. M., Roehrig, A. D., & Pressley, M. (2004). The first days of school in effective and less effective primary-grades classrooms. *Elementary School Journal, 104*, 269–287. doi:10.1086/499753
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, England: Wiley.
- *Bradshaw, C. P., Waasdorp, T. E., & Leaf, P. J. (2012). Effects of school-wide positive behavioral interventions and supports on child behavior problems. *Pediatrics, 130*, 1136–1145. doi:10.1542/peds.2012-0243

- Bradshaw, C. P., Zmuda, J. H., Kellam, S. G., & Lalongo, N. S. (2009). Longitudinal impact of two universal preventive interventions in first grade on educational outcomes in high school. *Journal of Educational Psychology, 101*, 926–937. doi:10.1037/a0016586
- *Brigman, G. A., & Webb, L. D. (2003). Ready to learn: Teaching kindergarten students school success skills. *Journal of Educational Research, 96*, 286–292. doi:10.1080/00220670309597641
- Brophy, J. (2006). History of research on classroom management. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 17–43). Mahwah, NJ: Lawrence Erlbaum.
- *Catalano, R. F., Mazza, J. J., Harachi, T. W., Abbott, R. D., Haggerty, K. P., & Fleming, C. B. (2003). Raising healthy children through enhancing social development in elementary school: Results after 1.5 years. *Journal of School Psychology, 41*, 143–164. doi:10.1016/S0022-4405(03)00031-1
- Chitiyo, M., May, M. E., & Chitiyo, G. (2012). An assessment of the evidence-base for School-Wide Positive Behavior Support. *Education & Treatment of Children, 35*, 1–24.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- *Connor, C. M., Ponitz, C. C., Phillips, B. M., Travis, Q. M., Glasney, S., & Morrison, F. J. (2010). First graders' literacy and self-regulation gains: The effect of individualizing student instruction. *Journal of School Psychology, 48*, 433–455. doi:10.1016/j.jsp.2010.06.003
- *Crean, H. F., & Johnson, D. B. (2013). Promoting Alternative Thinking Strategies (PATHS) and elementary school aged children's aggression: Results from a cluster randomized trial. *American Journal of Community Psychology, 52*, 56–72. doi:10.1007/s10464-013-9576-4
- *Curtis, C., & Norgate, R. (2007). An evaluation of the Promoting Alternative Thinking Strategies curriculum at Key Stage 1. *Educational Psychology in Practice, 23*, 33–44. doi:10.1080/02667360601154717
- *Daunic, A., Smith, S., Brank, E., & Penfield, R. (2006). Classroom-based cognitive-behavioral intervention to prevent aggression: Efficacy and social validity. *Journal of School Psychology, 44*, 123–139. doi:10.1016/j.jsp.2006.01.005
- Dignath, C., Buettner, G., & Langfeldt, H. (2008). How can primary school students learn self-regulated learning strategies most effectively? A meta-analysis on self-regulation training programmes. *Educational Research Review, 3*, 101–129. doi:10.1016/j.edurev.2008.02.003
- *Dion, E., Roux, C., Landry, D., Fuchs, D., Wehby, J., & Dupéré, V. (2011). Improving attention and preventing reading difficulties among low-income first-graders: A randomized study. *Prevention Science, 12*, 70–79. doi:10.1007/s11121-010-0182-5
- *Domitrovich, C. E., Cortes, R. C., & Greenberg, M. T. (2007). Improving young children's social and emotional competence: A randomized trial of the preschool "PATHS" curriculum. *Journal of Primary Prevention, 28*, 67–91. doi:10.1007/s10935-007-0081-0
- Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development, 82*, 405–432. doi:10.1111/j.1467-8624.2010.01564.x
- Erikson, E. H. (1968). *Identity: Youth and crisis*. New York, NY: W. W. Norton.

- Evertson, C. M., & Weinstein, C. S. (Eds.). (2006). *Handbook of classroom management: Research, practice, and contemporary issues*. Mahwah, NJ: Lawrence Erlbaum.
- Ferguson, C. J., Miguel, C. S., Kilburn, J. C., & Sanchez, P. (2007). The effectiveness of school-based anti-bullying programs: A meta-analytic review. *Criminal Justice Review, 32*, 401–414. doi:10.1177/0734016807311712
- *Flannery, D. J., Vazsonyi, A. T., Liau, A. K., Guo, S., Powell, K. E., Atha, H., . . . Embry, D. (2003). Initial behavior outcomes for the peacebuilders universal school-based violence prevention program. *Developmental Psychology, 39*, 292–308. doi:10.1037/0012-1649.39.2.292
- Flower, A., McKenna, J. W., Bunuan, R. L., Muething, C. S., & Vega, R., Jr. (2014). Effects of the Good Behavior Game on challenging behaviors in school settings. *Review of Educational Research, 84*, 546–571. doi:10.3102/0034654314536781
- Freiberg, H. J., & Lapointe, J. M. (2006). Research-based programs for preventing and solving discipline problems. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management. Research, practice, and contemporary issues* (pp. 735–786). New York, NY: Lawrence Erlbaum.
- *Frey, K. S., Nolen, S. B., Van Schoiack Edstrom, L., & Hirschstein, M. K. (2005). Effects of a school-based social-emotional competence program: Linking children's goals, attributions, and behavior. *Journal of Applied Developmental Psychology, 26*, 171–200. doi:10.1016/j.appdev.2004.12.002
- Froyen, L. A., & Iverson, A. M. (1999). *Schoolwide and classroom management: The reflective educator-leader*. Upper Saddle River, NJ: Prentice Hall.
- Greenberg, M. T., & Kusché, C. A. (1993). *Promoting social and emotional development in deaf children: The PATHS project*. Seattle: University of Washington Press.
- Greenberg, M. T., & Kusché, C. A. (2002). *Promoting Alternative Thinking Strategies: Blueprint for violence prevention* (2nd ed.). Boulder: Institute of Behavioral Sciences, University of Colorado.
- Grossman, D. C., Neckerman, H. J., Koepsell, T. D., Liu, P. Y., Asher, K. N., Beland, K., . . . Rivara, F. P. (1997). Effectiveness of a violence prevention curriculum among children in elementary school: A randomized controlled trial. *Journal of the American Medical Association, 277*, 1605–1611. doi:10.1001/jama.1997.03540440039030
- *Hamre, B. K., Pinata, R. C., Mashburn, A. J., & Downer, J. T. (2012). Promoting young children's social competence through the preschool PATHS curriculum and MyTeachingPartner professional development resources. *Early Education and Development, 23*, 809–832. doi:10.1080/10409289.2011.607360
- *Han, S. S., Catron, T., Weiss, B., & Marciel, K. K. (2005). A teacher-consultation approach to social skills training for pre-kindergarten children: Treatment model and short-term outcome effects. *Journal of Abnormal Child Psychology, 33*, 681–693. doi:10.1007/s10802-005-7647-1
- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London, England: Routledge.
- Hedges, L. V. (2007). Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics, 32*, 341–370. doi:10.3102/1076998606298043
- Hedges, L. V., & Hedberg, E. C. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis, 29*, 60–87. doi:10.3102/0162373707299706

- *Holen, S., Waaktaar, T., Lervåg, A., & Ystgaard, M. (2012). The effectiveness of a universal school-based programme on coping and mental health: A randomised, controlled study of Zippy's Friends. *Educational Psychology, 32*, 657–677. doi:10.1080/01443410.2012.686152
- *Holen, S., Waaktaar, T., Lervåg, A., & Ystgaard, M. (2013). Implementing a universal stress management program for young school children: Are there classroom climate or academic effects? *Scandinavian Journal of Educational Research, 57*, 420–444. doi:10.1080/00313831.2012.656320
- *Holsen, I., Smith, B. H., & Frey, K. S. (2008). Outcomes of the social competence program Second Step in Norwegian elementary schools. *School Psychology International, 29*, 71–88. doi:10.1177/0143034307088504
- *Horner, R. H., Sugai, G., Smolkowski, K., Eber, L., Nakasato, J., Todd, A. W., & Esperanza, J. (2009). A randomized, wait-list controlled effectiveness trial assessing School-Wide Positive Behavior Support in elementary schools. *Journal of Positive Behavior Interventions, 11*, 133–144. doi:10.1177/1098300709332067
- Jennings, J. L., & DiPrete, T. A. (2010). Teacher effects on social and behavioral skills in early elementary school. *Sociology of Education, 83*, 135–159. doi:10.1177/0038040710368011
- *Jones, S. M., Brown, J. L., & Lawrence Aber, J. (2011). Two-year impacts of a universal school-based social-emotional and literacy intervention: An experiment in translational developmental research. *Child Development, 82*, 533–554. doi:10.1111/j.1467-8624.2010.01560.x
- Jones, V. F., & Jones, L. S. (2012). *Comprehensive classroom management, creating communities of support and solving problems* (10th ed.). Upper Saddle River, NJ: Pearson.
- Kagan, S. (2005). *Structureel coöperatief leren* [Structural cooperative learning]. Vlissingen, Netherlands: Bazalt.
- Kärnä, A., Voeten, M., Little, T. D., Poskiparta, E., Kaljonen, A., & Salmivalli, C. (2011). A large-scale evaluation of the KiVa anti-bullying program. *Child Development, 82*, 311–330. doi:10.1111/j.1467-8624.2010.01557.x
- Kellam, S. G., Brown, C. H., Poduska, J., Lalongo, N., Wang, W., Toyinbo, P., . . . Wilcox, H. C. (2008). Effects of a universal classroom behavior management program in first and second grades on young adult behavioral, psychiatric, and social outcomes. *Drug and Alcohol Dependence, 95*(Suppl. 1), S5–S28. doi:10.1016/j.drugalcdep.2008.01.004
- Kelshaw-Levering, K., Sterling-Turner, H. E., & Henry, J. R. (2000). Randomized interdependent group contingencies: Group reinforcement with a twist. *Psychology in the Schools, 37*, 523–533. doi:10.1002/1520-6807(200011)37:63.0.CO;2-W
- *Koglin, U., & Petermann, F. (2011). The effectiveness of the behavioural training for preschool children. *European Early Childhood Education Research Journal, 19*, 97–111. doi:10.1080/1350293X.2011.548949
- *Kutnick, P., Ota, C., & Berdondini, L. (2008). Improving the effects of group working in classrooms with young school-aged children: Facilitating attainment, interaction and classroom activity. *Learning and Instruction, 18*, 83–95. doi:10.1016/j.learninstruc.2006.12.002
- Lane, K. L., Menzies, H. M., Bruhn, A. L., & Crnobori, M. (2011). *Managing challenging behaviors in schools: Research-based strategies that work*. New York, NY: Guilford Press.
- *Leflot, G., van Lier, P. A. C., Onghena, P., & Colpin, H. (2010). The role of teacher behavior management in the development of disruptive behaviors: An intervention

- study with the Good Behavior Game. *Journal of Abnormal Child Psychology*, 38, 869–882. doi:10.1007/s10802-010-9411-4
- Lewis, T. J., & Sugai, G. (1999). Effective behavior support: A systems approach to proactive schoolwide management. *Focus on Exceptional Children*, 31, 1–24.
- *Linares, L. O., Rosbruch, N., Stern, M. B., Edwards, M. E., Walker, G., Abikoff, H. B., & Alvir, J. (2005). Developing cognitive-social-emotional competencies to enhance academic learning. *Psychology in the Schools*, 42, 405–417. doi:0.1002/pits.20066
- *Louwe, J. J., Van Overveld, C. W., Merk, W., De Castro, B. O., & Kooops, W. (2007). De invloed van het programma alternatieve denkstrategieën op reactieve en proactieve agressie bij jongens in het primair onderwijs: Effecten na één jaar [The effects of the PATHS-curriculum on reactive and proactive aggression]. *Pedagogische Studiën*, 84, 277–292.
- *Lynch, K. B., Geller, S. R., & Schmidt, M. G. (2004). Multi-year evaluation of the effectiveness of a resilience-based prevention program for young children. *Journal of Primary Prevention*, 24, 335–353. doi:10.1023/B:JOPP.0000018052.12488.d1
- Lynch, S. A., & Simpson, C. G. (2010). Social skills: Laying the foundation for success. *Dimensions of Early Childhood*, 38(2), 3–12.
- Marzano, R. J., Marzano, J. S., & Pickering, D. J. (2003). *Classroom management that works: Research-based strategies for every teacher*. Alexandria, VA: Association for Supervision and Curriculum Development.
- *McClowry, S. G., Snow, D. L., Tamis-LeMonda, C., & Rodriguez, E. T. (2010). Testing the efficacy of INSIGHTS on student disruptive behavior, classroom management, and student competence in inner city primary grades. *School Mental Health*, 2, 23–35. doi:10.1007/s12310-009-9023-8
- Merrell, K. W., Gueldner, B. A., Ross, S. W., & Isava, D. M. (2008). How effective are school bullying intervention programs? A meta-analysis of intervention research. *School Psychology Quarterly*, 23, 26–42. doi:10.1037/1045-3830.23.1.26
- *Mishara, B., & Ystgaard, M. (2006). Effectiveness of a mental health promotion program to improve coping skills in young children: Zippy's Friends. *Early Childhood Research Quarterly*, 21, 110–123. doi:10.1016/j.ecresq.2006.01.002
- *Morris, P., Millenky, M., Raver, C. C., & Jones, S. M. (2013). Does a preschool social and emotional learning intervention pay off for classroom instruction and children's behavior and academic skills? Evidence from the foundations of learning project. *Early Education and Development*, 24, 1020–1042. doi:10.1080/10409289.2013.825187
- Oliver, R. M., Wehby, J. H., & Reschly, D. J. (2011). Teacher classroom management practices: Effects on disruptive or aggressive student behavior. *Campbell Systematic Reviews*, 4, 1–55. doi:10.4073/csr.2011.4
- O'Neill, S., & Stephenson, J. (2012). Does classroom management coursework influence pre-service teachers' perceived preparedness or confidence? *Teaching and Teacher Education*, 28, 1131–1143. doi:10.1016/j.tate.2012.06.008
- *Petermann, F., & Natzke, H. (2008). Preliminary results of a comprehensive approach to prevent antisocial behaviour in preschool and primary school pupils in Luxembourg. *School Psychology International*, 29, 606–626. doi:10.1177/0143034308099204
- Peters, J. H. (2012). Are they ready? Final year pre-service teachers' learning about managing student behaviour. *Australian Journal of Teacher Education*, 37, 18–42. doi:10.14221/ajte.2012v37n9.2
- Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2008). Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other

- causes of asymmetry. *Journal of Clinical Epidemiology*, 61, 991–996. doi:10.1016/j.jclinepi.2007.11.010
- Polanin, J. R. (2013). *Addressing the issue of meta-analysis multiplicity in education and psychology* (Unpublished doctoral dissertations). Loyola University Chicago, Chicago, IL. Retrieved from http://ecommons.luc.edu/luc_diss/539
- Raudenbush, S. W., Bryk, A. S., & Congdon, R. (2004). *HLM 6 for Windows* [Computer software]. Skokie, IL: Scientific Software International.
- *Raver, C. C., Jones, S. M., Li-Grining, C., Zhai, F., Metzger, M. W., & Solomon, B. (2009). Targeting children's behavior problems in preschool classrooms: A cluster-randomized control trial. *Journal of Consulting and Clinical Psychology*, 77, 302–316. doi:10.1037/a0015302
- Rhoades, B. L., Warren, H. K., Domitrovich, C. E., & Greenberg, M. T. (2011). Examining the link between preschool social-emotional competence and first grade academic achievement: The role of attention skills. *Early Childhood Research Quarterly*, 26, 182–191. doi:10.1016/j.ecresq.2010.07.003
- *Riggs, N. R., Greenberg, M. T., Kusché, C. A., & Pentz, M. A. (2006). The mediational role of neurocognition in the behavioral outcomes of a social-emotional prevention program in elementary school students: Effects of the PATHS curriculum. *Prevention Science*, 7, 91–102. doi:10.1007/s11121-005-0022-1
- *Rimm-Kaufman, S. E., Fan, X., Chiu, Y.-J., & You, W. (2007). The contribution of the Responsive Classroom Approach on children's academic achievement: Results from a three year longitudinal study. *Journal of School Psychology*, 45, 401–421. doi:10.1016/j.jsp.2006.10.003
- Rydell, A. M., & Henricsson, L. (2004). Elementary school teachers' strategies to handle externalizing classroom behaviour: A study of relations between perceived control, teacher orientation and strategy preferences. *Scandinavian Journal of Psychology*, 45, 93–102. doi:10.1111/j.1467-9450.2004.00384.x
- *Schick, A., & Cierpka, M. (2005). Faustlos: Evaluation of a curriculum to prevent violence in elementary schools. *Applied & Preventive Psychology*, 11, 157–165. doi:10.1016/j.appsy.2005.05.001
- Schussler, D. L., Poole, I. R., Whitlock, T. W., & Evertson, C. M. (2007). Layers and links: Learning to juggle “one more thing” in the classroom. *Teaching and Teacher Education*, 23, 572–585. doi:10.1016/j.tate.2007.01.016
- *Seifer, R., Gouley, K., Miller, A. L., & Zakriski, A. (2004). Implementation of the PATHS curriculum in an urban elementary school. *Early Education and Development*, 15, 471–485. doi:10.1207/s15566935eed1504_6
- *Self-Brown, S., & Mathews, S. (2003). Effects of classroom structure on student achievement goal orientation. *Journal of Educational Research*, 97, 106–111. doi:10.1080/00220670309597513
- Shook, A. C. (2012). A study of preservice educators' dispositions to change behavior management strategies. *Preventing School Failure*, 56, 129–136. doi:10.1080/1045988X.2011.606440
- Smart, J. B., & Brent, I. L. (2010). A grounded theory of behavior management strategy selection, implementation, and perceived effectiveness reported by first-year elementary teachers. *Elementary School Journal*, 110, 567–584. doi:10.1086/651196
- *Snyder, J., Low, S., Schultz, T., Barner, S., Moreno, D., Garst, M., . . . Schrepferman, L. (2011). The impact of brief teacher training on classroom management and child behavior in at-risk preschool settings: Mediators and treatment utility. *Journal of Applied Developmental Psychology*, 32, 336–346. doi:10.1016/j.appdev.2011.06.001

- *Sørli, M., & Ogden, T. (2007). Immediate impacts of PALS: A school-wide multi-level programme targeting behaviour problems in elementary school. *Scandinavian Journal of Educational Research*, 51, 471–492. doi:10.1080/00313830701576581
- Swinson, J., Woof, C., & Melling, R. (2003). Including emotional and behavioural difficulties pupils in a mainstream comprehensive: A study of the behaviour of pupils and classes. *Educational Psychology in Practice*, 19, 65–75. doi:10.1080/0266736032000061215
- *Tiano, J. D., & McNeil, C. B. (2006). Training Head Start teachers in behavior management using parent-child interaction therapy: A preliminary investigation. *Journal of Early and Intensive Behavior Intervention*, 3, 220–233.
- Ttofi, M. M., & Farrington, D. P. (2011). Effectiveness of school-based programs to reduce bullying: A systematic and meta-analytic review. *Journal of Experimental Criminology*, 7, 27–56. doi:10.1007/s11292-010-9109-1
- Van de Grift, W., Van der Wal, M., & Torenbeek, M. (2011). Ontwikkeling in de pedagogisch didactische vaardigheid van leraren in het basisonderwijs [Development in the pedagogical and didactic skills of primary school teachers]. *Pedagogische Studiën*, 88, 416–432.
- *Van Lier, P. A. C., Vuijk, P., & Crijnen, A. A. M. (2005). Understanding mechanisms of change in the development of antisocial behavior: The impact of a universal intervention. *Journal of Abnormal Child Psychology*, 33, 521–535. doi:10.1007/s10802-005-6735-7
- *Vespo, J. E., Capece, D., & Behforooz, B. (2006). Effects of the nurturing curriculum on social, emotional, and academic behaviors in kindergarten classrooms. *Journal of Research in Childhood Education*, 20, 275–285.
- *Webster-Stratton, C., Reid, M. J., & Stoolmiller, M. (2008). Preventing conduct problems and improving school readiness: Evaluation of the Incredible Years Teacher and Child Training Programs in high-risk schools. *Journal of Child Psychology and Psychiatry*, 49, 471–488. doi:10.1111/j.1469-7610.2007.01861.x
- Weissberg, R. P., Resnik, H., Payton, J., & O'Brien, M. U. (2003). Evaluating social and emotional learning programs. *Educational Leadership*, 60, 46–50.
- What Works Clearinghouse. (2014). *Procedures and standards handbook* (Version 3.0). Retrieved from http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v3_0_draft_standards_handbook.pdf
- *Witvliet, M., van Lier, P. A. C., Cuijpers, P., & Koot, H. M. (2009). Testing links between childhood positive peer relations and externalizing outcomes through a randomized controlled intervention study. *Journal of Consulting and Clinical Psychology*, 77, 905–915. doi:10.1037/a0014597
- Woodcock, S., & Reupert, A. (2012). A cross-sectional study of student teachers' behaviour management strategies throughout their training years. *Australian Educational Researcher*, 39, 159–172. doi:10.1007/s13384-012-0056-x
- Wubbels, T., Brekelmans, M., Van Tartwijk, J., & Admiraal, W. (1999). Interpersonal relationships between teachers and students in the classroom. In H. C. Waxman & H. J. Walberg (Eds.), *New directions for teaching practice and research* (pp. 151–170). Berkeley, CA: McCutchan.
- Wubbels, T., Den Brok, P., Veldman, I., & Van Tartwijk, J. (2006). Leraar-leerlingrelaties en klassenmanagement in het hedendaagse onderwijs [Teacher-student relationships and classroom management in present-day education]. In R. Klarus & W. Wardekker (Eds.), *Wat is goed onderwijs: Bijdragen vanuit de pedagogiek* [What is good education: Contributions from the pedagogy] (pp. 121–141). Den Haag, Netherlands: Boom Lemma Uitgevers.

Authors

HANKE KORPERSHOEK is an assistant professor at the Groningen Institute for Educational Research, University of Groningen, Grote Rozenstraat 3, 9712 TG Groningen, the Netherlands; e-mail: h.korpershoek@rug.nl. Her research focuses on gender differences in career choices among adolescents, classroom management, and the impact of student motivation and school commitment on educational outcomes.

TRUUS HARMS is a senior researcher at the Groningen Institute for Educational Research, University of Groningen, Grote Rozenstraat 3, 9712 TG Groningen, the Netherlands; e-mail: g.j.harms@rug.nl. Her research regards innovations in (vocational) education and their impact on teachers' roles and students' motivation and behavior.

HESTER DE BOER is a postdoctoral researcher at the Groningen Institute for Educational Research, University of Groningen, Grote Rozenstraat 3, 9712 TG Groningen, the 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.

MECHTELD VAN KUIJK is a researcher and lecturer at the Groningen Institute for Educational Research, University of Groningen, Grote Rozenstraat 3, 9712 TG Groningen, the Netherlands; e-mail: m.f.van.kuijk@rug.nl. Her research interests include teacher professional development, data use, reading comprehension, effective instruction, goal setting, and educational effectiveness.

SIMONE DOOLAARD is an assistant professor at the Groningen Institute for Educational Research, University of Groningen, Grote Rozenstraat 3, 9712 TG Groningen, the Netherlands; e-mail: s.doolaard@rug.nl. Her research focuses in general on educational effectiveness in primary education and on teachers and teaching behavior.