

Reducing disruptive behaviours and improving classroom behavioural climate with class-wide positive behaviour support in middle schools

Vesa Närhi^{a,b,*}, Tiina Kiiski^c and Hannu Savolainen^a

^aUniversity of Jyväskylä, Jyväskylä, Finland; ^bUniversity of Eastern Finland, Joensuu, Finland; ^cKoskela Middle School, Mänttä-Vilppula, Finland

Disruptive behaviour in classrooms is a significant challenge for learning in schools and a risk factor for students' academic achievement and a significant source of teachers' work-related stress. Earlier research shows that clear behavioural expectations, monitoring students' adherence to them and behaviour-specific praise are effective practices to reduce disruptive behaviour. Although behaviour problems are common in middle schools, most of the interventions have been developed and studied in elementary schools. This randomised study evaluated the effects of a class-wide intervention on classroom behavioural climate and disruptive behaviour, on teacher-experienced stress and on the time needed for behaviour management in middle school. The classes were selected for intervention by their teachers on the basis of poor behavioural climate. The intervention was based on teachers' cooperation; they collectively agreed on clear behavioural expectations, used positive feedback and, if needed, applied consequences in response to high rates of disruptive behaviour. The results indicated medium to large effects on classroom behavioural climate according to teachers' evaluations, and somewhat more inconsistent effects on classroom behavioural climate according to student evaluations and in the time needed for behaviour management. The behavioural climate of the classes remained at a constant level during the follow-up. The intervention was well accepted by teachers and students. The results suggest that an easily applicable intervention may produce significant improvements in classroom behavioural climate in middle schools.

Keywords: universal prevention; consultation; disruptive behaviour; class-wide intervention

Introduction

Disruptive behaviour in a school classroom can have many negative impacts for both teachers and students. Problem behaviours are clear risks for students' academic achievement at school (e.g. Frick *et al.*, 1991; Wagner *et al.*, 2005) and for post-school adjustment (e.g. Schaeffer *et al.*, 2006; Karakus *et al.*, 2012). For teachers, students' problem behaviours are a cause for increased work-related stress and reduced wellbeing (Boyle *et al.*, 1995; Klassen & Chiu, 2010). Behaviour problems are more common in middle schools than in elementary schools (Kaufman *et al.*, 2009), yet most of the interventions and studies are conducted in elementary schools.

*Corresponding author. Faculty of Education and Psychology, University of Jyväskylä, PO Box 35, 40014 University of Jyväskylä, Finland E-mail: vesa.m.narhi@jyu.fi.

New interventions are needed to address the needs of middle schools to reduce problem behaviour and to improve the behavioural climate in classrooms (see Emmer & Gerwels, 2013). This study was conducted to evaluate the effects of a classroom intervention based on clear behavioural expectations, immediate positive feedback and weekly consequences for middle-school classes with poor behavioural climate.

Classroom management refers to the actions teachers take to create environments supporting academic and social/emotional learning (Evertson & Weinstein, 2013). Recent meta-analyses (Oliver *et al.*, 2011; Korpershoek *et al.*, 2016) have agreed that teachers' systematic classroom management practices reduce problem behaviours in classrooms, but, largely due to the variability of the interventions studied and the variability of the measures used, no firm conclusions on the most effective components of interventions have been reached with meta-analytic methods. Two systematic reviews have attempted to identify the critical components of classroom management. Simonsen *et al.* (2008) identified five categories of evidence-supported practices: maximising structure and predictability; providing clear behavioural expectations and feedback on meeting these expectations; actively engaging students; acknowledging appropriate behaviour; and responding to inappropriate behaviour. Epstein *et al.* (2008) concluded similar, but somewhat differently described, categories of classroom practices as being evidence-based: identifying the antecedents and consequences of problem behaviour and tailoring interventions based on the observations; modifying the classroom environment (e.g. by revisiting and reinforcing behavioural expectations); and teaching and reinforcing appropriate behaviour. A third systematic review on teacher-led interventions (Whear *et al.*, 2013) concluded that positive feedback was a common element of programmes reporting significant results.

Most of the studies included in the systematic reviews and meta-analysis were conducted in elementary schools. Few class-wide interventions have been studied in middle schools. Gottfredson *et al.* (1993) reported an improvement in student behaviour with an intervention that included multiple components at various levels. At school level the school discipline policy was revised, school-parent cooperation was enhanced by increasing communication with parents and classroom practices were targeted by improving classroom organisation, clarifying behavioural expectations, reinforcing appropriate behaviour and consistently responding to misbehaviour. Similarly, with a multicomponent classroom intervention (including seating arrangements, reminders about classroom rules, teacher's evaluative comments on students' behaviour at the end of the lessons, arranging behaviour-dependent, valued activities for the students and students' self-evaluations on following the rules), an increase in on-task behaviour was reported by McNamara *et al.* (1986). In neither of the studies did the design allow for analysis of the independent effects of different components on outcome variables. Johnson *et al.* (1996) compared the effects of three class-wide interventions—class syllabus combined with individual student achievement assessment, self-monitoring and active teaching of classroom expectations with behaviour-specific praise—in a middle-school class. Of the interventions, teaching classroom expectations with behaviour-specific praise was the most effective and, when applied by different teachers, also resulted in a decrease in disruptive behaviour during lessons.

The few studies conducted in middle schools and showing a decrease in disruptive behaviour had at least two common elements, clear behavioural expectations of the students and positive feedback for meeting the expectations. This was also the most effective intervention in Johnson *et al.*'s (1996) study. Clear expectations and feedback were also identified as evidence-based methods by Simonsen *et al.* (2008) and Epstein *et al.* (2008). Thus, it seems that they are the core of successful classroom management, and also applicable to middle schools.

The middle-school environment is challenging both for students and for implementing interventions. Each class of students is instructed by several teachers, each of them having at least partly individual expectations for students' behaviour and procedures of classroom management. For students, this creates an environment which poses high demands for behaviour regulation, as they have to adapt to expectations varying from lesson to lesson. This aspect of middle schools has not been fully discussed in earlier interventions, despite its obvious importance for classroom behavioural climate.

Setting common behavioural expectations at school level is a central feature of the School-Wide Positive Behaviour Intervention and Support (SWPBIS; Warren *et al.*, 2006) approach. In SWBPIS, universal behavioural support is provided to all students and the assumption is that the whole school staff is committed to the expectations and reinforcing them, primarily with positive feedback. The aim is to make the school a more predictable environment for the students, and thus support appropriate behaviour. There is ample evidence that having clear expectations is crucial in reducing disruptive behaviour at the school level (Horner *et al.*, 2010), and thus we applied this principle of having commonly agreed behavioural expectations at lesson level as a critical part of the intervention.

In this intervention study, clarifying and unifying behavioural expectations during lessons, offering immediate positive feedback for students and applying weekly consequences if failing to meet the expectations were designed to be the core components of the intervention. To enhance the teachers' willingness to carry out the intervention, we attempted to secure the critical issues of the acceptability of the intervention identified by Elliott (1988); namely, minimising the time needed for implementing the intervention and ensuring that the teachers experienced behavioural problems of the classes so severe that intervention was necessary. We also attempted to guarantee administrative support (see Han & Weiss, 2005) by contacting the schools via the principals, and common commitment from teachers by instructing principals to make the decision on participation in a staff meeting.

A pilot study (Närhi *et al.*, 2014) using similar interventions with 28 middle-school classes found that during the intervention, the learning climate improved (according to both teachers' and students' evaluations), the time needed for behaviour management and teachers' experienced strain decreased, and the intervention was highly acceptable. However, as no control group was included, the magnitude of the change resulting directly from the intervention could not be estimated. In the pilot study the intervention was carried out in schools with the support of an external consultant. As the aim was to develop an intervention the schools could carry out with their own personnel, no external consultants were used in the current study. We also used a more stringent randomised delayed control group design.

Methods

The study was conducted as part of a larger research project funded by the Finnish Ministry of Education and Culture. Ethical approval was obtained from the Committee of Research Ethics of the University of Eastern Finland.

Setting

Finnish middle school is compulsory for all children. It includes the last three years of comprehensive school (i.e. grades seven to nine), with students from 12 to 16 years of age. Finland is in general committed to inclusive educational principles, although separate special classes for students with severe learning disabilities and psychiatric and behaviour problems still exist. No official tracking of students on the basis of achievement level is used. The teaching is based on a subject-teacher system, so each group of students is taught by 8 to 12 teachers. The school year in middle schools is divided into periods lasting from six to nine weeks (varying by municipality). Consequently, the teachers teaching each class vary from one period to another. Each class of students is designated a homeroom teacher, who meets students every week and is responsible, for example, for overall monitoring of how students manage in school and informing them on concurrent issues and functioning as a liaison between school and parents.

Participants

The principals of middle schools were approached with a letter instructing them to introduce and discuss the project at a staff meeting and to make the decision to participate on the basis of the staff-meeting response. They were asked to select classes in which several of the teachers considered the classroom behavioural climate to be poor. The maximum number of classes to participate within each school was two, to reduce the workload for the teachers and to secure the possibility of carrying out the intervention with good fidelity. After classes were selected, the parents of the students were approached with a letter, sent from the schools, describing the intervention and the study, and asking them to sign their consent for their children to participate in the study.

Altogether, the principals of all 175 schools within the area of the study were approached with an introductory letter. Thirty-eight schools (21.7%) were willing to participate, of which three special schools were excluded from the study. Not having any information on the non-interested schools, we were not able to analyse the representativeness of the study sample.

The schools selected only seventh-grade (58%) and eighth-grade (42%) classes for the study. The mean number of students in a class was 19.2 (SD = 3.5; range 13 to 27), the mean proportion of males was 54.5% (SD = 9.5; range 29.4% to 76.9%), the mean proportion of students with special educational status was 5.7% (SD = 8.8; range 0% to 36%) and the mean proportion of students having research consent from their parents was 82.8% (SD = 18.9; range 36.0 to 100).

In total, 607 teachers participated in the assessments, and all background information was received from 577 of them.

Intervention

Each school nominated, according to its own needs, one to four participants for training on carrying out the consultation for teachers. The training consisted of two days prior to and one half-day during the implementation of the intervention. During the first two days the trainees were provided with the intervention manual (Kiiski *et al.*, 2012) and familiarised with the intervention and consultation procedures. The third day was arranged one to two weeks after the beginning of the intervention; it was consultation-based and focused on questions and troubleshooting in the implementation. In each school the intervention started with an approximately three-hour launch meeting. The meeting consisted of basic information on classroom management, especially on the importance of clearly articulated behavioural expectations and positive feedback for appropriate behaviour. In the meeting the teachers discussed and agreed on the two most important student behaviours that disrupted the classroom behavioural climate of the participating class. The teachers were then instructed to rephrase the disruptive behaviours as behavioural instructions for the students (e.g. from 'The students talk out of turn' to 'I wait for my turn to talk'). In the meeting the teachers also set a weekly minimum target for students' success in meeting the goals. The target was generally set at 70% to 80%.

The teachers were instructed to give behaviour-specific praise during and after the lessons when the students succeeded in following the behavioural instructions. Teachers were guided to give the praise in a simple and concrete manner, telling the students exactly what they had done right (e.g. 'Thank you for waiting your turn to answer'). After the launch meeting the homeroom teacher explained the intervention, the reasons for it and its goals to his/her class. The subject teachers were instructed to discuss the intervention with their students and to specify what the behavioural goals meant during their lessons. During the intervention the teachers evaluated the behaviour of each student in relation to the goals on a yes or no basis. The evaluations were conducted at the end of each lesson, for which three to five minutes were reserved. Teachers were instructed to discuss the evaluations with the students and to focus on the students who had succeeded in behaving according to expectations. The teachers returned their evaluation sheets at the end of each week to the homeroom teacher, who summarised the evaluations for each student (taking altogether 20 to 30 minutes). The homeroom teachers were instructed to use a portion of a weekly homeroom session for giving feedback to the students based on the weekly summaries at the whole-class level. The homeroom teachers were instructed to focus on the positive issues in students' behaviour.

If a student did not achieve the weekly target, the homeroom teacher informed the student services team, which contacted the student's parents. When a student did not achieve the target for the first time, a discussion with him/her and his/her parents was arranged at school within the next week. If a student failed to achieve the target again, discussions with a team of teachers and, possibly, other school personnel (appointed by the school) were arranged. The discussions during these meetings were restricted to the behavioural expectations set for the class and carried out according to a future-oriented and supportive scheme described in the manual. This consequence was

selected as it corresponds well with the instructions for Finnish schools on addressing disruptive behaviour.

A special education teacher from each school was nominated to deliver a written synopsis of the intervention and instruct new teachers at each periodic change. The student services team informed the parents about the intervention and the reasons for it with a letter; some schools also arranged a parent meeting relating to the intervention.

During the intervention a teacher meeting, led by the staff member trained, was arranged at four to five-week intervals. At the meetings each teacher was asked his/her opinions about the intervention, and the importance of clear expectations and behaviour-specific praise for reducing disruptive behaviour was further elaborated. The teachers also made a consensus decision regarding the continuation of the intervention.

Measures

The changes in the classroom behavioural climate during the intervention were assessed using questionnaires for students and teachers. We used the definition of discipline problems in the classroom by Levin and Nolan (2010) as the basis for classroom behavioural climate assessment, and defined a good classroom behavioural climate as a lack of discipline problems. Based on this definition, a good classroom behavioural climate exists when (1) students' behaviour is geared towards learning, (2) students do not compromise the rights of others to learn, (3) the learning situation is psychologically and physically safe and (4) students take appropriate care of the physical classroom environment. The questionnaires were developed, tested and modified in a preparatory study (Peitso *et al.*, 2011) and the same used in a pilot study (Närhi *et al.*, 2014). Online questionnaires were used for both teachers and students. The links to the questionnaires were sent to the teachers based on lists provided by the school administrations. The links to the student questionnaires were sent to the homeroom teacher, who arranged the assessment at school.

The teacher questionnaire consisted of 17 statements on a six-point Likert scale (asking them to evaluate if the statement described working at the classroom level from 'very poorly' to 'very well'). Four of the statements were related to students' concentration on learning (e.g. 'Students pay attention to teaching well'; range of Cronbach's α at different measurement points = .88 to .91), five to disruptive behaviours (e.g. 'It is too noisy during lessons'; α = .88 to .92), five to physical and psychological safety (e.g. 'Students mock each other for answering incorrectly'; α = .75 to .79) and three to caring for the classroom environment (e.g. 'Students leave the classroom tidy after the lessons'; α = .78 to .82). With the same scale, the teachers also answered four questions evaluating the strain they experienced in teaching the participating class (e.g. 'I feel stressed about teaching this class'; α = .87 to .88). The teachers also approximated the average proportion of a lesson they used for behaviour management with the class by marking it on a scale with 5% intervals.

The student questionnaire on the classroom behavioural climate consisted of 21 statements on a four-point Likert scale (ranging from 'In my classroom this happens never' to 'In my classroom this happens in all lessons'). Six of the statements were

taken from the Finnish PISA questionnaire (Holopainen *et al.*, 2009). Two of these statements were also used to compare the classes with the Finnish PISA sample. The reliabilities were similar to those in the pilot study and consistently good for the scales ‘Concentrating on learning’ ($\alpha = .80$ to $.82$) and ‘Disruptive behaviour’ ($\alpha = .83$ to $.85$) and moderate for the scales ‘Safety’ ($\alpha = .66$ to $.73$) and ‘Caring for classroom environment’ ($\alpha = .63$ to $.73$).

The teachers and the students evaluated the acceptability of the intervention at post-intervention assessments with a six-point Likert scale. The statements (e.g. ‘The intervention does not have negative effects for the students’) were adapted and modified from the Intervention Rating Profile 15 questionnaire (Martens *et al.*, 1985). The teachers’ acceptability measure consisted of eight items, the scale of α being $.94$, and the students’ measure consisted of seven items, with α of $.92$.

Treatment fidelity was assessed at different levels; all fidelity measures were conducted for the study to include the critical elements of the intervention. *The consultants* filled in a yes/no checklist on completed phases (described in the manual) of the launch and follow-up meetings; a percentage of completed phases was used as an index of fidelity. A six-point Likert scale (totally disagree to totally agree) was used for all other fidelity measures. *The teachers* evaluated the fidelity of the start-up of the intervention at the school level (e.g. ‘Starting the intervention was agreed in a staff meeting’, Teacher Start-Up Fidelity; 5 items; scale $\alpha = .86$). The teachers who did not participate in the launch meeting (either because of inconvenience or the fact that they only taught the class during later periods) evaluated the quality of instructing new teachers (e.g. ‘The behavioural expectations of the class were clear to me after familiarisation’, Teacher Familiarisation Fidelity; 5 items; scale $\alpha = .89$). Teachers also evaluated their own performance according to the principles of the intervention in relation to the instructions and feedback (e.g. ‘I gave students verbal feedback on how they had succeeded with the behavioural goals’, Teacher Feedback Fidelity; 3 items; scale $\alpha = .84$) and in relation to the technical aspects (e.g. ‘I filled in the evaluation sheet of the classes’ lessons’, Teacher Technical Fidelity; 3 items; scale $\alpha = .78$) of the intervention. *The students* evaluated the teachers’ behaviour in relation to the principles of the intervention (e.g. ‘Teachers gave positive feedback to those students who succeeded with the behavioural expectations’, Students’ Evaluation on Fidelity; 10 items; scale $\alpha = .91$).

Design and statistical procedures

Based on the development of the intervention and experiences from the pilot studies (Peitso *et al.*, 2011; Närhi *et al.*, 2014), we aimed at a two-month intervention. To allow time for pre- and post-measurements and to avoid having the winter break in the middle of the intervention, we chose to carry out the intervention in the autumn term for the first group and in the spring term for the second group (the schools were randomised into two groups with a random number generator). The randomisation was carried out at the school level to avoid the spill-over effect of the intervention. The randomisation produced 15 schools (25 classes) for the experimental group (EG) and 20 schools (29 classes) for the delayed control group (DcG). Because the

schools needed to know the timetable for the study to plan for the school year, the randomisation was carried out prior to the first measurement.

The intervention for EG was carried out during the autumn term, and for DcG during the spring term of the school year 2013–14. During the school year there were four measurements: the first before the intervention for EG; the second during the last month of the autumn term (post-intervention measurement for EG); the third before the intervention for DcG; and the fourth during the last month of the spring term (post-intervention measurement for DcG). The follow-up measurement was conducted in the middle of the autumn term of the next school year. In line with the implementation nature of the research, the teachers made decisions on the continuation or termination of the intervention. For most of the classes the intervention was active at post-intervention measurement (measurement 2 for EG, 95.8% and measurement 4 for DcG, 82.6% with no difference between the groups: $\chi^2(1) = 2.161$; $p = .142$). The duration of the intervention, from introducing it to the students to the post-intervention measurement, did not differ between groups [independent samples t -test: $t(41,065) = 0.519$; $p = .607$; for EG, $M = 51.3$ days, $SD = 10.3$; for DcG, $M = 52.7$ days, $SD = 7.2$].

Owing to misunderstandings and scheduling problems in 64% of the EG classes, the majority of the teachers completed the first measurement only after the intervention launch meeting. Apparently the launch meeting had an effect on teachers' evaluations, as teachers evaluated the classroom behavioural climate of these classes as having been systematically better than classes evaluated before the launch meeting, the effect sizes being medium to large (on classroom behavioural climate scales, all $p < .035$ in independent samples t -tests). No significant differences were observed in students' evaluations (on classroom behavioural climate scales, all $p > .26$ in independent samples t -tests). This suggests that the launch meeting itself had an effect on teachers' evaluations, although the intervention itself had not started.

The participation of the classes and available teacher and student evaluations at different measurement points is described in Figure 1. Four classes, all from DcG, withdrew from the study after the autumn term and prior to their intervention, one due to school work load, one because the behavioural climate had improved and the teachers deemed the intervention unnecessary; the reason for the withdrawal of the other two classes remained unknown. These classes were dropped from the main analysis but were included in the additional analysis on the intervention effects during the autumn term. In addition, one class from each group did not participate in the third student assessment, so the number of classes is 25 in each group for the analysis on teacher evaluations and 24 in each group for student evaluations.

The number of teachers participating in the measurement showed a somewhat decreasing trend during the study. The mean numbers of teachers evaluating each class were 11.0 (3.2), 10.7 (3.4), 9.0 (3.1), 9.6 (3.2) and 8.9 (2.8) from the first to the follow-up assessment, respectively. As we did not have reliable information on the number of teachers teaching each class at different points of the school year, we were not able to detect if the trend was due to curriculum changes or due to teachers being unwilling to participate. The students' participation remained at a more constant level during the study, although the mean percentages of students from each class

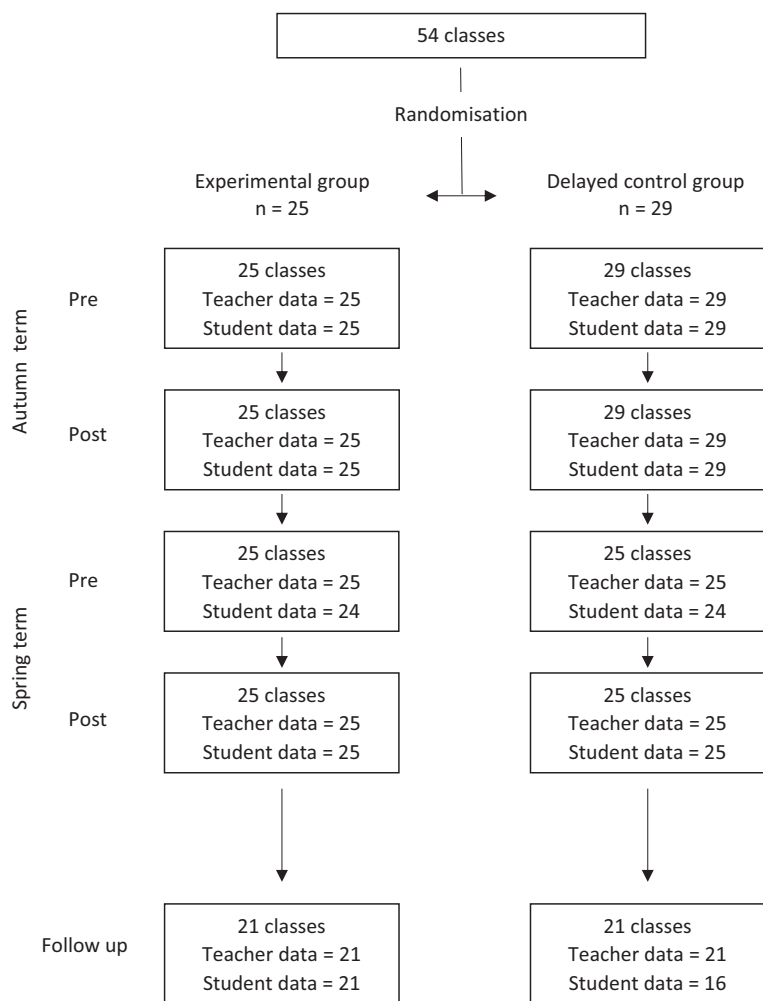


Figure 1. Flowchart of the classes participating the study and the number of classes of which the teachers' and students' evaluations were available at different measurement points

participating in the measurements also decreased slightly, being 78.5 (19.9), 75.4 (20.2), 74.3 (18.6), 67.6 (19.3) and 67.3 (22.7) from the first to the follow-up measurement, respectively.

The changes in classroom behavioural climate were evaluated through repeated analysis of variance (ANOVA) measures with repeated contrasts. Along with the research design, the interest was in group by measurement time interactions between the pre- and post-measurements of the autumn and spring terms, as the intervention was carried out between those points, for EG and DcG, respectively. Since there was a semester change and a winter break of approximately two weeks between the post-measurement of the autumn and the pre-measurement of the spring term, we decided to report possible effects between these two measurements too. The significance level was set at $p < .05$. In the ANOVA the effect size was measured with partial eta squared ($p\eta^2$) and interpreted using the guidelines of Cohen (1988), where

the thresholds for small, medium and large effect sizes are 0.01, 0.06 and 0.14, respectively. For group comparisons with *t*-tests, Cohen's *d* (computed as the difference between group means divided by the pooled standard deviation) was used as a measure of effect size, with standard thresholds for small, medium and large effect sizes of 0.20, 0.50 and 0.80, respectively. Since the intervention was targeted at the whole classroom, the changes were evaluated at the classroom level by aggregating the teachers' and students' evaluations separately for each class at each measurement phase. The analyses are similar to those used in the pilot study (Närhi *et al.*, 2014), to allow comparison of the results. To analyse the stability of the change, the follow-up measurement at the next autumn term was compared with the last measurement in the spring with repeated ANOVA measures.

Results

The classes were nominated for the study by the teachers on the basis of poor behavioural climate. For student evaluations of the behavioural climate, the study sample (at pre-intervention measurement in the autumn) was compared with the Finnish PISA 2009 sample, which was a representative sample of Finnish 15-year-old students in middle schools (OECD, 2009). The students of this study evaluated their possibilities of concentrating on learning (i.e. 'Students cannot work well') in their classroom to be significantly poorer than students in the PISA sample, but with negligible effect size [$M = 2.07$, $SD = 0.59$ and $M = 1.94$, $SD = 0.78$, respectively; independent samples *t*-test, $t(1229.45) = 5.48$; $p < .001$; Cohen's $d = 0.17$]. No significant differences were observed on disruptive behaviour ['There is noise and disorder'; present sample $M = 2.48$, $SD = 0.74$, PISA sample $M = 2.50$, $SD = 0.82$; independent samples *t*-test, $t(1087.67) = 0.488$; $p = .325$; Cohen's $d = 0.02$]. This indicates that the behavioural climate at baseline measurement in the participating classes was not perceived especially poorly by the student participants of this study.

Table 1. Descriptive results of the fidelity measures and results of independent samples *t*-tests

Fidelity measure	Experimental group ($n = 25$)		Delayed control group ($n = 25$)		$t(48)$	p	d^*
	M	SD	M	SD			
Consultant: Launch meeting (%)	95.67 ^a	5.65	97.45 ^b	3.81	1.268 ^c	.212	0.36
Consultant: Follow-up meetings (%)	90.04 ^c	13.14	89.02 ^d	14.92	0.244 ^f	.809	0.08
Teacher: Start-up	5.02	0.25	4.93	0.30	1.118	.269	0.35
Teacher: Familiarisation ^g	4.61	1.12	4.68	0.52	0.271 ^h	.789	0.08
Teacher: Feedback	4.46	0.36	4.38	0.30	0.878	.384	0.24
Teacher: Technical	4.80	0.43	4.69	0.52	0.781	.469	0.23
Students	4.16 ^a	0.44	4.01 ^a	0.41	1.218 ⁱ	.230	0.35

*Cohen's *d*.

^a $n = 24$. ^b $n = 22$. ^c $n = 21$. ^d $n = 24$. ^eCorrected d.f. = 40.581. ^fd.f. = 43. ^g $n = 21$ in both G1 and G2. ^hCorrected d.f. = 28.166. ⁱd.f. = 46.

Table 2. Descriptive statistics of the teacher evaluations and results of the repeated-measures ANOVAs from the pre-measurement in the autumn term to the post-measurement in the spring term

Scale	Experimental group (<i>n</i> = 25)						Delayed control group (<i>n</i> = 25)					
	Autumn			Spring			Follow-up ^d			Autumn		
	Pre <i>M</i> (SD)	Post <i>M</i> (SD)		Pre <i>M</i> (SD)	Post <i>M</i> (SD)		<i>M</i> (SD)			Pre <i>M</i> (SD)	Post <i>M</i> (SD)	Follow-up ^d <i>M</i> (SD)
Concentrating on learning ^{a, b, c}	3.97 (0.53)	4.17 (0.50)		4.15 (0.43)	4.07 (0.48)		4.19 (0.48)			3.62 (0.49)	3.79 (0.46)	4.14 (0.46)
Disruptive behaviour ^{a, b, c}	3.44 (0.52)	3.11 (0.56)		3.12 (0.50)	3.15 (0.45)		3.06 (0.43)			3.70 (0.53)	3.52 (0.44)	3.09 (0.44)
Safety ^b	4.62 (0.34)	4.70 (0.41)		4.69 (0.43)	4.63 (0.38)		4.73 (0.36)			4.45 (0.47)	4.50 (0.39)	4.80 (0.40)
Caring for classroom ^{b, c}	4.62 (0.35)	4.63 (0.38)		4.57 (0.28)	4.59 (0.34)		4.67 (0.41)			4.39 (0.41)	4.34 (0.38)	4.62 (0.34)
For behaviour management (%) ^{a, b, c}	17.57 (6.92)	16.15 (7.40)		14.92 (5.42)	14.62 (5.05)		16.20 (4.43)			23.60 (7.26)	20.36 (6.12)	15.14 (4.99)
Strain ^b	2.29 (0.45)	2.26 (0.47)		2.28 (0.44)	2.29 (0.42)		2.08 (0.40)			2.56 (0.53)	2.47 (0.47)	2.14 (0.36)

Note: Experimental group had the intervention between pre- and post-measurements in the autumn term. Delayed control group had the intervention between pre- and post-measurements in the spring term.

^aMain effect for time from the pre-measurement in the autumn to the post-measurement in the spring significant at *p* < .05.

^bGroup by time interaction from the pre-measurement in the autumn to the post-measurement in the spring significant at *p* < .05.

^cMain effect for group from the pre-measurement in the autumn to the post-measurement in the spring significant at *p* < .05. Experimental group performing better than delayed control group.

^d*n* = 21.

Table 3. Results of the repeated-measures contrast of group by time interactions of the ANOVA analyses of the teacher evaluations

	Autumn; pre- vs. post-measurement			Post-measurement autumn vs. pre-measurement spring			Spring; pre- vs. post-measurement		
	<i>F</i> (1, 48)	<i>p</i>	<i>p</i> η^2	<i>F</i> (1, 48)	<i>p</i>	<i>p</i> η^2	<i>F</i> (1, 48)	<i>p</i>	<i>p</i> η^2
Concentrating on learning	4.638	.036	.088	4.213	.046	.081	14.210	<.001	.228
Disruptive behaviour	16.309	<.001	.254	5.789	.020	.108	23.351	<.001	.327
Safety	1.000	.322	.020	1.219	.275	.025	6.629	.016	.116
Caring for classroom	1.226	.274	.025	1.314	.257	.027	4.106	.048	.079
Proportion of lesson for BM	0.544	.465	.011	1.430	.238	.029	6.413	.015	.118
Strain	1.371	.247	.028	3.593	.064	.070	3.397	.071	.068

Note: *p* η^2 = partial eta squared.

Fidelity

The results on the fidelity measures are presented in Table 1. Comparisons indicated that there were no significant differences between the groups on any fidelity measure, in terms of effect sizes the differences were negligible to small, so it seemed that the intervention was carried out with equal fidelity in both groups. Based on their self-evaluations, the consultants followed the manual closely. Teachers' self-evaluations on the fidelity of the intervention indicated rather high fidelity, the means being close to five on a scale of one to six. The highest fidelity ratings were obtained on scales describing the 'technical' aspects of the intervention, while the ratings on the interactions with students according to the principles of the intervention were somewhat lower. The lowest fidelity ratings, although still well on the positive side of the scale, were obtained from the students, who evaluated their teachers' behaviour according to the principles of the intervention.

Teacher evaluations

As there were no differences between groups on the background variables or on the fidelity measures, the subsequent analyses on teacher and student evaluations were conducted without covariates. The descriptive statistics of the teacher evaluations are presented in Table 2, and the results of the repeated-measures ANOVA contrasts in Table 3.

The intervention had a partial effect on the teacher-evaluated behavioural climate in the classrooms. During the autumn semester, the behavioural climate improved more in EG than in DcG on two scales, 'Concentrating on learning' and 'Disruptive behaviour'. The effect size was large on 'Disruptive behaviour' and moderate for 'Concentrating on learning'.¹ The intervention effects in the spring term were clearer, as larger effect sizes across several outcome scales indicated. More specifically, in the

Table 4. Descriptive statistics of the student evaluations and results of the repeated-measures ANOVA analyses from the pre-measurement at the autumn term to the post-measurement at the spring term

Scale	Experimental group (<i>n</i> = 24)						Delayed control group (<i>n</i> = 24)					
	Autumn			Spring			Follow-up ^d			Autumn		
	Pre <i>M</i> (SD)	Post <i>M</i> (SD)	Pre <i>M</i> (SD)	Post <i>M</i> (SD)	Pre <i>M</i> (SD)	Post <i>M</i> (SD)	Pre <i>M</i> (SD)	Post <i>M</i> (SD)	Pre <i>M</i> (SD)	Post <i>M</i> (SD)	Pre <i>M</i> (SD)	Post <i>M</i> (SD)
Concentrating on learning ^{a, b, c}	2.78 (0.21)	2.88 (0.23)	2.90 (0.21)	2.89 (0.20)	2.90 (0.21)	2.89 (0.20)	2.95 (0.23)	2.58 (0.27)	2.63 (0.27)	2.71 (0.27)	2.83 (0.22)	2.87 (0.28)
Disruptive behaviour ^{a, b, c}	2.22 (0.28)	2.11 (0.27)	2.03 (0.24)	2.04 (0.23)	2.03 (0.24)	2.04 (0.23)	1.96 (0.24)	2.46 (0.29)	2.38 (0.27)	2.27 (0.27)	2.13 (0.19)	2.05 (0.28)
Safety	3.24 (0.19)	3.21 (0.26)	3.26 (0.27)	3.21 (0.27)	3.26 (0.27)	3.21 (0.27)	3.26 (0.30)	3.14 (0.23)	3.11 (0.23)	3.14 (0.22)	3.16 (0.26)	3.26 (0.12)
Caring for classroom ^{b, c}	3.27 (0.24)	3.24 (0.26)	3.19 (0.26)	3.10 (0.34)	3.19 (0.26)	3.10 (0.34)	3.24 (0.31)	3.16 (0.22)	3.09 (0.20)	3.01 (0.25)	3.11 (0.24)	3.12 (0.28)

Note: Experimental group had the intervention between pre- and post-assessments in the autumn term. Delayed control group had the intervention between pre- and post-assessments in the spring term.

^aMain effect for time from the pre-measurement in the autumn to the post-measurement in the spring significant at *p* < .05.

^bGroup by time interaction from the pre-measurement in the autumn to the post-measurement in the spring significant at *p* < .05.

^cMain effect for group from the pre-measurement in the autumn to the post-measurement in the spring significant at *p* < .05. Experimental group performing better than delayed control group.

^d*n* = 21.

^e*n* = 16.

Table 5. Results of the repeated-measures contrast of group by time interactions of the ANOVA analysis of student evaluations

	Autumn; pre- vs. post-measurement			Post-measurement autumn vs. pre-measurement spring			Spring; pre- vs. post-measurement		
	$F(1, 46)$	p	$p\eta^2$	$F(1, 46)$	p	$p\eta^2$	$F(1, 46)$	p	$p\eta^2$
Concentrating on learning	0.689	.411	.015	1.908	.174	.040	8.426	.006	.155
Disruptive behaviour	0.400	.530	.009	0.602	.442	.013	5.312	.026	.104
Safety	0.001	.972	.000	0.177	.676	.004	1.186	.282	.025
Caring for classroom	0.712	.403	.015	0.264	.610	.006	7.812	.008	.145

Note: $p\eta^2$ = partial eta squared.

spring the intervention effect was significant with moderate effect sizes also for the scales of 'Safety' and 'Caring for classroom environment' and for the time teachers reported to have used for behaviour management. The group by time interaction was not significant on the strain experienced in teaching the class, although the effect size was in the moderate range. Unexpectedly, the group by time interaction was also significant between the post-measurement in autumn and the pre-measurement in spring (including change of semester and winter break in school) on 'Concentrating on learning' and 'Disruptive behaviour'.

The stability of the intervention effects during the school year was analysed with EG with repeated-measures ANOVAs on 'Concentrating on learning' and 'Disruptive behaviour' at autumn post-, spring pre- and spring post-measurements. The results indicated that the teacher-evaluated classroom behavioural climate stayed at a constant level over the school year, as the within-subject models were non-significant for both measures, $F(1.501, 36.020) = 1.239$; $p = .299$ and $F(1.588, 38.102) = 0.163$; $p = .800$, respectively.

The follow-up data on the following school year in the middle of the autumn semester on teacher evaluations was available for 21 classes in both groups. The results of the ANOVA analysis indicated no changes on any of the classroom behavioural climate scales, on strain experienced in teaching the class or on time used for behaviour management between the post-measurement in spring and follow-up (for all main effects of time $p_{\min} = .097$), also all main effects for group ($p_{\min} = .185$) and group by time interactions ($p_{\min} = .163$) were non-significant.

Student evaluations

The descriptive statistics for the student evaluations of the scales are presented in Table 4, and the results of the repeated-measures ANOVA contrasts in Table 5.

There were no intervention effects on the classroom behavioural climate according to student evaluations during the autumn intervention.² During the spring term, the intervention effects were significant, with medium to large effect sizes on scales of 'Concentration on learning', 'Disruptive behaviour' and 'Caring for classroom environment'. Unlike in the teacher evaluations, there were no significant group by time

interaction effects between the post-measurement in autumn and the pre-measurement in spring (i.e. over the winter break).

The follow-up data on student evaluations was available for 21 classes in EG and 16 classes in DcG. The results of the ANOVA analysis on classroom behavioural climate scales indicated a significant reduction in disruptive behaviour from post-measurement in spring to follow-up [the main effect of time; $F(1, 35) = 5.512$; $p = .025$; $\eta^2 = .136$]. For all other learning climate scales, the main effects of time were non-significant ($p_{\min} = .074$). Also, all main effects for group ($p_{\min} = .215$) and group by within-subject interactions ($p_{\min} = .215$) were non-significant (Table 4).

Intervention acceptability

There were no differences between groups on either teacher [EG, $M = 4.32$, $SD = 0.48$; DcG, $M = 4.40$, $SD = 0.40$; independent samples t -test, $t(48) = 0.654$, $p = .516$] or student [EG, $M = 3.92$, $SD = 0.58$; DcG, $M = 3.85$, $SD = 0.54$; independent samples t -test, $t(48) = 0.425$, $p = .673$] evaluations on intervention acceptability. In general, both seemed to evaluate the intervention as acceptable, as the evaluations were clearly on the positive side of the scale. The teachers seemed to evaluate the intervention as more acceptable than the students, but the direct comparability of the measures is questionable.

Discussion

In the study, the effects of a classroom-wide intervention on the classroom behavioural climate were studied in middle schools. We used a randomised design, in which the intervention was implemented during the autumn term for one group and in the spring term for the other group of classes. The classes were selected by the teachers on the basis of poor behavioural climate. The intervention was carried out by the teachers supported by monthly meetings, which were directed by a trained member of the school staff. The changes in the classroom behavioural climate were evaluated by teacher and student evaluations. The teachers' evaluations of the classroom behavioural climate showed an improvement during the intervention, the improvement being larger and more consistent for the DcG having intervention in the spring term. The students' evaluations showed intervention-related improvement only during intervention in the spring term. There were no intervention-related reductions in teacher-reported strain of teaching the classes, although the effect in the spring term approached significance with a moderate effect size. The classroom behavioural climate stayed at a constant level from the post-intervention measure through the spring term for the classes that participated in the intervention in the autumn, and the follow-up indicated no deterioration in the behavioural climate next autumn.

The evidence on the effects of interventions for improving classroom behavioural climate is largely based on studies conducted in elementary schools. The results of this study indicate that an intervention based on the effective principles identified mostly in elementary-school studies, namely clearly articulated behavioural expectations and positive feedback for appropriate behaviour, is effective also in the middle school. Oliver *et al.* (2011), in their meta-analysis, reported a group-level effect size

of 0.80 (with 95% CI 0.51 to 1.09) on the teacher-evaluated disruptive behaviour at the post-measurement of the studies. In the present study, the comparable effect size at the autumn intervention post-measurement was 0.99. This indicates that the intervention produced similar effects to interventions studied mostly in elementary schools. This is also notable in the light of the little amount of training required to implement the intervention. The intervention was guided by school personnel, who were trained for the consultations for 2 days in groups. Within each school, teachers were instructed on the intervention in a three-hour launch meeting, followed by monthly one-hour meetings. This amount of training is considerably low, in comparison with that required for earlier-reported teacher-led interventions (see Whear *et al.*, 2013).

In the intervention, the teachers teaching each class agreed on common behavioural expectations for all lessons. This is a feature which has not been highlighted in earlier middle-school interventions, but which we deem to be important considering the large number of teachers teaching each class. With the common behavioural expectations, the lessons are more predictable for the students, and that reduces their need to adjust their behaviour to individual teacher expectations from lesson to lesson. Based on anecdotal evidence from the teachers, having common behavioural expectations encourages teachers to apply them systematically with all classes and reduces their need to repeatedly rationalise the expectations for the students.

Different results for the autumn and spring-term interventions were not expected. The groups did not differ in any measures of the fidelity, also the background variables of the groups were similar. It also seems unlikely that the intervention effect itself is term-dependent. One possible reason is the baseline difference between the groups. The behavioural climate in the classes which participated the intervention during the autumn term was systematically better at baseline. The most likely reason for the difference is the flaw in the timing of the assessments, as for two-thirds of the EG classes the majority of teachers responded to the baseline questionnaire only after the launch meeting, and this appeared to have had a systematic positive effect on teachers', but not on students', evaluations. Another possibility is that, for some unknown reason, the randomisation produced groups in which there were real differences in behavioural climates. However, it did not seem that the flaw in the timing could inflate the intervention effects. Another unexpected observation was the significant improvement in two teacher scales in DcG, in relation to EG, over the winter holiday. One possible explanation is that in DcG there was more room for improvement, and the effects of the winter holidays were observed in that group but not in EG.

One major limitation in intervention research in general is the lack of follow-up. The intervention during the autumn resulted in improvement in teacher evaluations on classroom behavioural climate, and these evaluations stayed at a constant level over the spring term. Although the follow-up data next school year were not available from all classes, it also showed no deterioration of behavioural climate. Based on these observations, it seems that the intervention produced lasting improvements in classroom behavioural climate. One reason for the sustained effect could be that the intervention is not only about supporting students' behaviour but also, to a smaller or

greater extent, about changing the working culture within a school. Here the most likely critical new element is that the teachers share their views on educational principles, in this case about behavioural expectations, and agree on a common strategy to improve the situation. Finnish schools have high autonomy, extending also to individual teachers' pedagogical approaches, which has many positive implications but may also produce negative effects, in particular with regard to variability in behavioural expectations and inconsistency in providing feedback for students.

While the delay in the baseline assessment is a serious limitation for the study, it also suggests, quite interestingly, that mere knowledge that there is a way to improve classroom behavioural climate and agreeing with colleagues on the ways to work with the class may have positive effects for teachers' feelings towards classroom behavioural climate. If this were so, it suggests that just increased collaboration among middle-school teachers on responding to student disruptive behaviour may have a positive effect on teachers' perceptions of the students.

Interesting comparisons can also be made with the results of the pilot study (Närhi *et al.*, 2014), as the classroom behavioural climate measures and recruiting procedures were identical in the studies. Yet, the results of the pilot study clearly indicated a larger improvement in the learning climates of the classes during the intervention than in either group of the present study. The most obvious is the difference between the samples of the studies. The classes, which participated in the pilot study, were on average clearly poorer on all classroom behavioural climate measures than the classes participating in this study at baseline, so part of the results of the pilot study may be explained by the regression towards the mean effect. On the contrary, the observation tentatively suggests reserving this intervention for classes with significant problems in behavioural climate.

The interventions in the studies were highly similar, in terms of actual implementation identical, but there were two differences we would like to discuss here (although their possible effects on the differing results cannot be estimated with the present data). The first is that in the present study, in line with the aim to study the intervention implemented by schools with minimal external support, the consultants in the teacher meetings were schools' own personnel, while in Närhi *et al.*'s (2014) study external consultants were used. While it might be tempting to suggest that the skill levels of consultants in studies were different and that might explain the differences, in the pilot study there were no differences in results between more and less-experienced consultants (Närhi *et al.*, 2014). The consultants (mainly special education teachers) in the present study were carefully trained, the consultation meetings were guided by the manual and based on their self-evaluations they adhered to the manual closely. However, it is still possible that, as each consultant in the pilot study consulted several schools, they gained more experience than in the present study. Another possible reason that could have affected the difference in results is that teachers may find external consultants more acceptable or convincing than a consultant belonging to the school staff.

The other, and only, difference in the actual interventions between the studies was that the criterion for delivering consequences for high-level disruptive behaviour of a student was defined in the present study as 'not meeting the behavioural goals at target level', while in the pilot study it was defined as 'exceeding the acceptable limit of

disruptive behaviour'. This change was made on the basis of teacher feedback from the pilot study, and to decrease the possible punitive aspects of intervention. The final point from the combined results of the studies is that it is possible that these kinds of interventions are in fact more effective during the spring than the autumn term, as the pilot study was conducted in the spring term, during which the effects were also more favourable in the present study.

There are also limitations which should be considered when interpreting results. The most clear was the wrongly timed pre-measurement for EG, discussed earlier. Second, of the invited schools only one-fifth participated, and not having information on the non-participating schools, we are not able to conclude on the representativeness of the sample included in the study. Third, the number of teachers participating in the measurements per class showed a somewhat decreasing trend, and this may have produced unreliability in the results. Fourth, due to the requirements set by the participating schools, the randomisation had to be conducted prior to the pre-assessment, so the teachers were aware of the timing of the intervention. This could have had an effect on the baseline evaluations, the direction and amount of which remains unknown. Fifth, while we measured several aspects of the fidelity in carrying out the intervention, the fidelity measures were based on consultant, teacher and student evaluations only, and no observational measures were used, which would have produced a more reliable estimate of the fidelity. The same reservation and suggestion applies also to the classroom behavioural climate measures used. Finally, only about 80% of the classes participated in the follow-up measurement, and the number of teachers and the proportion of students per class were somewhat lower, so the results concerning the follow-up should be considered preliminary.

Conclusions

The results suggest that with an easily teachable and applicable intervention, the classroom behavioural climate in middle-school classes can be enhanced. The intervention has been shown to be acceptable for both teachers and students. The intervention is based on clear behavioural expectations and frequent and immediate feedback for students on meeting these expectations; these elements have been consistently identified as important in classroom management. A key element of the intervention is teachers working in cooperation, which we deem important in itself in enhancing the working and learning climate in middle schools. In future it would be essential to study what are the most important implementation, school, teacher or student-related factors that moderate the effects of this kind of intervention.

NOTES

¹ When the intervention effects during the autumn term were analysed with repeated-measures ANOVAs, including the four classes in DeG which dropped out of the study after the autumn term post-measurement, the group by time interactions were significant for 'Concentrating on learning' [$F(1, 52) = 4.166; p = .046, p\eta^2 = .074$] and 'Disruptive behaviour' [$F(1, 52) = 14.093; p < .001, p\eta^2 = .213$]. These results indicate that the significant results observed in the analysis of the whole design were not due to classes dropping out of the study.

² When the intervention effects during the autumn term were analysed with repeated-measures ANOVAs, including the four classes in DeG which dropped out of the study after the second assessment, no group by assessment interactions were significant (for the four scales, $p_{\min} = .429$), a result which corresponds with the reported analysis.

References

- Boyle, G. J., Borg, M. G., Falzon, J. M. & Baglioni, A. J. (1995) A structural model of the dimensions of teacher stress, *British Journal of Educational Psychology*, 65(1), 49–67.
- Cohen, J. (1988) *Statistical power analysis for the behavioral sciences* (2nd edn) (Hillsdale, NJ, Lawrence Erlbaum Associates).
- Elliott, S. N. (1988) Acceptability of behavioral treatments: Review of variables that influence treatment selection, *Professional Psychology: Research and Practice*, 19(1), 68–80.
- Emmer, E. T. & Gerwels, M. C. (2013) Classroom management in middle and high school classrooms, in: E. Emmer, E. Sabornie, C. M. Evertson & C. S. Weinstein (Eds) *Handbook of classroom management: Research, practice, and contemporary issues* (New York, NY, Routledge), 407–437.
- Epstein, M., Atkins, M., Cullinan, D., Kutash, K. & Weaver, K. (2008) *Reducing behavior problems in the elementary school classroom* (Washington, D.C., National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, US Department of Education).
- Evertson, C. M. & Weinstein, C. S. (2013) Classroom management as a field of inquiry, in: E. Emmer, E. Sabornie, C. M. Evertson & C. S. Weinstein (Eds) *Handbook of classroom management: Research, practice, and contemporary issues* (New York, NY, Routledge), 3–15.
- Frick, P. J., Kamphaus, R. W., Lahey, B. B., Loeber, R., Christ, M. A. G., Hart, E. L. & Tannenbaum, L. E. (1991) Academic underachievement and the disruptive behavior disorders, *Journal of Consulting and Clinical Psychology*, 59(2), 289–294.
- Gottfredson, D. C., Gottfredson, G. D. & Hybl, L. G. (1993) Managing adolescent behavior: A multiyear, multischool study, *American Educational Research Journal*, 30(1), 179–215.
- Han, S. S. & Weiss, B. (2005) Sustainability of teacher implementation of school-based mental health programs, *Journal of Abnormal Child Psychology*, 33(6), 665–679.
- Holopainen, P., Järvinen, R., Kuusela, J. & Packalen, P. (2009) *Työrauha tavaksi. Kohtaaminen, toimintakulttuuri ja pedagogiikka koulun arjessa* (Helsinki, Opetushallitus).
- Horner, R., Sugai, G. & Anderson, C. M. (2010) Examining the evidence base for school-wide positive behavior support, *Focus on Exceptional Children*, 42(1), 1–14.
- Johnson, T. C., Stoner, G. & Green, S. K. (1996) Demonstrating the experimenting society model with classwide behavior management interventions, *School Psychology Review*, 25, 199–214.
- Karakus, M. C., Salkever, D. S., Slade, E. P., Ialongo, N. & Stuart, E. (2012) Implications of middle school behavior problems for high school graduation and employment outcomes of young adults: Estimation of a recursive model, *Education Economics*, 20(1), 33–52.
- Kaufman, J. S., Jaser, S. S., Vaughan, E. L., Reynolds, J. S., Di Donato, J., Bernard, S. N. & Hernandez-Brereton, M. (2009) Patterns in office referral data by grade, race/ethnicity, and gender, *Journal of Positive Behavior Interventions*, 12(1), 44–54.
- Kiiski, T., Närhi, V. & Peitso, S. (2012) *KUMMI 9. Työrauha kaikille. Toimintamalli työrauhaongelmien vähentämiseksi* (Jyväskylä, Niilo Mäki Instituutti).
- Klassen, R. M. & Chiu, M. M. (2010) Effects on teachers' self-efficacy and job satisfaction: Teacher gender, years of experience, and job stress, *Journal of Educational Psychology*, 102(3), 741–756.
- Korpershoek, H., Harms, T., de Boer, H., van Kuijk, M. & Doolaard, S. (2016) A meta-analysis of the effects of classroom management strategies and classroom management programs on students' academic, behavioral, emotional, and motivational outcomes, *Review of Educational Research*, 86(3), 643–680.
- Levin, J. & Nolan, J. F. (2010) *Principles of Classroom management: A professional decision-making Model*, (6th edn) (Boston, MA, Allyn & Bacon).

- Martens, B. K., Witt, J. C., Elliott, S. N. & Darveaux, D. X. (1985) Teacher judgments concerning the acceptability of school-based interventions, *Professional Psychology: Research and Practice*, 16 (2), 191–198.
- McNamara, E., Evans, M. & Hill, W. (1986) The reduction of disruptive behaviour in two secondary school classes, *British Journal of Educational Psychology*, 56(2), 209–215.
- Närhi, V., Kiiski, T., Peitso, S. & Savolainen, H. (2014) Reducing disruptive behaviours and improving learning climates with class-wide positive behaviour support in middle schools, *European Journal of Special Needs Education*, 30(2), 274–285.
- OECD (2009) *OECD Programme for International Student Assessment (PISA)*. Available online at: pisa2009.acer.edu.au/downloads.php (accessed 5 November 2011).
- Oliver, R. M., Wehby, J. H. & Reschly, D. J. (2011) Teacher classroom management practices: Effects on disruptive or aggressive student behavior, *Campbell Systematic Reviews*, 2011, 4. <https://doi.org/10.4073/csr.2011.4>.
- Peitso, S., Kiiski, T. & Närhi, V. (2011) *Pilottitutkimus Työrauha kaikille -mallin vaikuttavuudesta*. Oppimisvaikeuksien tutkimus Suomessa, 14. valtakunnallinen tutkijatapaaminen, 19–20/5/2011, Helsinki, Finland.
- Schaeffer, C. M., Petras, H., Ialongo, N., Masyn, K. E., Hubbard, S., Poduska, J. & Kellam, S. (2006) A comparison of girls' and boys' aggressive-disruptive behavior trajectories across elementary school: Prediction to young adult antisocial outcomes, *Journal of Consulting and Clinical Psychology*, 74(3), 500–510.
- Simonsen, B., Fairbanks, S., Briesch, A., Myers, D. & Sugai, G. (2008) Evidence-based practices in classroom management: Considerations for research to practice, *Education and Treatment of Children*, 31(3), 351–380.
- Wagner, M., Kutash, K., Duchnowski, A. J., Epstein, M. H. & Sumi, W. C. (2005) The children and youth we serve: A national picture of the characteristics of students with emotional disturbances receiving special education, *Journal of Emotional and Behavioral Disorders*, 13(2), 79–96.
- Warren, J. S., Bohanon-Edmonson, H. M., Turnbull, A. P., Sailor, W., Wickham, D., Griggs, P. & Beech, S. E. (2006) School-wide positive behavior support: Addressing behavior problems that impede student learning, *Educational Psychology Review*, 18(2), 187–198.
- Whear, R., Thompson-Coon, J., Boddy, K., Ford, T., Racey, D. & Stein, K. (2013) The effect of teacher-led interventions on social and emotional behaviour in primary school children: A systematic review, *British Educational Research Journal*, 39(2), 383–420.