



A randomized controlled trial of the impact of a teacher classroom management program on the classroom behavior of children with and without behavior problems

Judy Hutchings^{a,*}, Pam Martin-Forbes^a, David Daley^b, Margiad Elen Williams^a

^a School of Psychology, Bangor University, Gwynedd LL57 2PZ, UK

^b Division of Psychiatry, Queens Medical Centre, University of Nottingham, NG7 2UH, UK

ARTICLE INFO

Article history:

Received 26 December 2011

Received in revised form 1 August 2013

Accepted 9 August 2013

Keywords:

Classroom intervention

Child behavior problems

Teachers

School-based

ABSTRACT

This randomized controlled trial (RCT) evaluated the efficacy of the Incredible Years (IY) Teacher Classroom Management (TCM; Webster-Stratton & Reid, 2002) program to assess whether training teachers in IY-TCM principles improve teacher behavior, whether any observed improvements impact pupil behavior classroom-wide, and whether these effects can be demonstrated with children at risk of developing conduct problems. Six intervention and six control classrooms comprising 12 teachers and 107 children (aged 3 to 7 years) were recruited. Children were screened for high or low behavior problems using the cut-off points of the teacher-rated Strengths and Difficulties Questionnaire (Goodman, 1997). The primary outcome measure was independent classroom observations using the Teacher–Pupil Observation Tool (Martin et al., 2010). Multilevel modeling analyses were conducted to examine the effect of the intervention on teacher, classroom, and child behavior. Results showed a significant reduction in classroom off-task behavior ($d = 0.53$), teacher negatives to target children ($d = 0.36$), target child negatives towards the teacher ($d = 0.42$), and target child off-task behavior ($d = 0.48$). These preliminary results demonstrate the potential impact of IY-TCM on both teacher and child behavior.

© 2013 Society for the Study of School Psychology. Published by Elsevier Ltd. All rights reserved.

1. Introduction

In the UK between three and seven percent of children aged five to 15 years meet diagnostic criteria for conduct disorder (CD; National Institute for Health Clinical Excellence [NICE], 2006); boys are three times more likely than girls to have such problems (Hutchings, Williams, Martin, & Pritchard, 2011; Office for National Statistics, 2007). Children with early onset behavioral problems likely to develop into CD are at high risk for social and emotional problems, poor school attendance, school dropout, academic failure and delinquency (Webster-Stratton, Reid, & Stoolmiller, 2008). Over the last decade, teachers have reported increasing levels of behavioral problems within the classroom (Hutchings et al., 2011). These children are often taught by teachers who are ill prepared to cope with disruptive behavior (Webster-Stratton et al., 2008). They are also likely to receive less support and positive feedback from their teachers and their peers (Arnold et al., 1999). Exposure to a supportive teacher and a positive classroom environment improves the academic achievement of high-risk children (Werner, 1999). High levels of praise for appropriate behavior improve children's behavioral, social, and emotional adjustment as do the use of proactive teaching and

* Corresponding author at: Centre for Evidence Based Early Intervention, Nantlle Building, Normal Site, Bangor University, Bangor, Gwynedd LL57 2PZ, UK. Tel.: +44 01248 383625.

E-mail address: j.hutchings@bangor.ac.uk (J. Hutchings).

ACTION EDITOR: John Carlson.

positive discipline strategies (Webster-Stratton et al., 2008). These studies demonstrate that there is a need for effective, evidence-based classroom intervention programs to support teachers.

1.1. The incredible years (IY) teacher classroom management program

The IY series is a suite of evidence-based programs (Webster-Stratton, 2005) for parents of children aged 0 to 12 years, therapeutic and classroom-based programs for children, and a Teacher Classroom Management (TCM) program. The series has demonstrated efficacy and effectiveness through high-quality randomized controlled trials (RCTs) and is one of 11 “Blueprints” for Violence Prevention Model Programs (Webster-Stratton et al., 2001a, Webster-Stratton et al., 2001b) identified from over 900 programs reviewed.

The IY-TCM program (Webster-Stratton & Reid, 2002) is delivered to groups of teachers for one day a month for five or six months. It aims to improve teacher–pupil relationships by improving home–school links, increasing teacher competencies in supporting children in the classroom, and developing children's social and problem-solving skills. Specific skills include proactive teacher strategies around rules and transitions and ensuring that fewer instructions are given but that they are clear specific and positive (i.e., that they tell the child what to do rather than what not to do; Webster-Stratton & Reid, 2002; Webster-Stratton & Taylor, 2001). It also incorporates all of the strategies that maximize behavior change through a collaborative delivery style (Hutchings, Gardner, & Lane, 2004; Webster-Stratton & Herbert, 1994). Teachers identify key classroom management skills through discussion, observation of videotaped examples of classroom situations, role-play rehearsal and classroom-based practice between sessions, including behavior planning. Feedback is provided at the start of the following session, and verbal and written assignments are reviewed and returned (Webster-Stratton & Reid, 2002).

1.2. Supporting evidence for IY-TCM program

A number of RCTs of the IY series have included the TCM program, in various combinations with the parent and child programs. Outcomes vary between studies with some measuring both teacher and child outcomes, while others have focused on only teacher or only child outcomes.

1.2.1. Teacher and child outcomes

The first trial was conducted by Webster-Stratton et al. (2001b) and examined the effectiveness of the IY parent and teacher training programs with 4-year-old children and their mothers. Participants were randomly assigned to either the intervention condition, which included both the IY parent and teacher training, or a control condition. Children in the intervention condition showed significantly less conduct problems at school, and teachers in the intervention condition showed significantly better classroom management skills than children and teachers in the control condition. No effect sizes were reported in this trial.

A second trial by Webster-Stratton, Reid, and Hammond (2004) with children diagnosed with oppositional defiant disorder included five combinations of IY programs (three of which included the TCM program) and a no-treatment control group. Conditions that included the TCM program showed significant effects on teacher's negative behavior compared with controls, with Cohen's *d* values ranging from 0.46 to 0.63 depending on the condition. There was also a significant reduction in children's negative behavior in the conditions including TCM training (*ds* = 0.41–0.46).

A third trial of the IY-TCM program conducted by Webster-Stratton et al. (2008) examined the effects of the program in combination with the IY child training component. Participants were teachers and children from Head Start, kindergarten, and first-grade classrooms. Multilevel modeling was used to assess the intervention effects and point-wise effect sizes were calculated as the intervention slope mean divided by the standard deviation of the teacher level normative slope. They found that teachers in the intervention condition showed significantly fewer harsh and critical strategies ($z = -0.49$) compared to teachers in the control classrooms. Children in intervention classrooms showed a significant reduction in conduct problems ($z = 0.03$) and levels of disengagement ($z = -0.14$).

Shernoff and Kratochwill (2007) examined the IY-TCM program in conjunction with mental health consultation. The aims of their study were to examine the transportability of the program to a preschool setting and to investigate the potential contextual barriers of implementation. They compared two methods of training, namely videotape modeling (VM) versus videotape modeling plus consultation (VMC). Teachers in the VMC condition reported significantly higher confidence ratings, significantly higher acceptability ratings, and significantly greater use of proactive instructional practices. Preschoolers in both conditions displayed nonsignificant reductions in disruptive behaviors; however, preschoolers in the VMC condition showed additional positive trends in social competence and increased adaptation to the school environment. No effect sizes were reported. Time was considered the most significant barrier to completing the training, while implementation barriers included coteacher's lack of exposure to program content and lack of congruence regarding appropriate classroom management strategies.

Williford and Shelton (2008) also conducted a study examining the effectiveness of an adaptation of the IY-TCM program delivered using mental health consultation. Ninety-six preschoolers who displayed disruptive behaviors, their teachers, and their caregivers participated in the study. Teachers in the intervention group reported significantly greater use of effective teaching strategies and stable levels of child disruptive behaviors across the school year compared to teachers in the control group who reported increasing levels of disruptive behavior across the school year. Greater ease of implementation and usefulness of the teaching strategies was also reported by teachers in the intervention group. No effect sizes were reported.

Raver et al. (2008) again examined an adapted version of the IY-TCM program with mental health consultation. This adapted version was part of the Chicago School Readiness Project that aimed to improve teachers' emotionally supportive classroom practices. Head Start teachers were randomly allocated to an intervention condition including the IY-TCM adaptation or a control group. Classrooms in the intervention condition showed significant improvements in emotional climate ($d = 0.89$). Teachers were more enthusiastic and more responsive to their students' needs ($d = 0.53$), and had less use of harsh and emotionally negative discipline practices ($d = 0.64$), however these findings were nonsignificant. Teachers in the intervention group were also more likely to display improved classroom management practices and better skills in monitoring and preventing student disruptive behavior in proactive ways ($d = 0.52$). Again, these differences were nonsignificant.

1.2.2. Child outcomes only

In addition to a pilot study (Baker-Henningham, Walker, Powell, & Gardner, 2009), Baker-Henningham, Scott, Jones, and Walker (2012) conducted a cluster RCT of the IY-TCM program in 24 Jamaican preschools. Effect sizes were calculated by dividing the regression coefficient with the baseline standard deviation. They found significant benefits in child behavior based on observations, teacher reports and parent reports. There were significant reductions in observed conduct problems ($r = .42$), supported by a significant decrease in teacher-reported behavior difficulties ($r = .47$) and parent-reported behavior difficulties ($r = .22$), and a significant increase in observed friendship skills ($r = .74$), supported by a significant increase in teacher-reported social skills ($r = .59$).

Herman, Borden, Reinke, and Webster-Stratton (2011) used data from the Webster-Stratton et al. (2004) trial to examine the effectiveness of the IY series with children with co-occurring internalizing problems. They found that children in the conditions including the IY-TCM program showed no significant reductions in internalizing symptom scores (from pretest). There were also no significant differences found at one year follow-up.

1.2.3. Teacher outcomes only

Carlson, Tiret, Bender, and Benson (2011) examined preschool teachers' perceptions of the usefulness and frequency of use of classroom management strategies following completion of the IY-TCM training. Twenty-four preschool teachers attended 100% of sessions (eight sessions and 32 h of training). No effect sizes were reported. Significant increases in teachers' perceptions of positive strategy use were found from pretest to posttest. Changes in teachers' perceptions were also maintained 16 weeks later.

1.3. Rationale for study

All of the studies discussed have shown promising results for the IY-TCM program; however, all were either implementing an adapted version of the program or alongside another intervention, so it was not possible to isolate the effectiveness of the TCM program. The IY-TCM program is potentially an effective stand-alone program for primary school teachers because every trained teacher is in contact with up to 30 children a year. This study evaluates the program as a stand-alone intervention. The study aims to assess whether the IY-TCM program reduces negative teacher and pupil behaviors and increases positive behaviors using an RCT design and independent, blind observation as the main outcome measure. Based on previous research (e.g., Webster-Stratton et al., 2001b; Webster-Stratton et al., 2004), the first hypothesis was that teachers in the intervention group would demonstrate postintervention fewer negative behaviors and more positive behaviors towards the whole class than at preintervention. The second hypothesis was that children in the intervention group would show fewer negative and more positive behaviors than their control counterparts. The third hypothesis was that teachers in the intervention group would show more positive and fewer negative behaviors towards children with high levels of behavior problems. Lastly, it was hypothesized that children with high levels of behavior problems in the intervention group would show greater behavioral changes from preintervention to postintervention compared to children with high levels of behavior problems in the control group.

2. Method

2.1. Participants

Teachers from 12 classes from 11 primary schools participated in the study (see Table 1). Eight classes were multiyear classes consisting of children ages 3 to 7, and four classes were single age reception classes (ages 4 to 5). Class sizes ranged from 14 to 30 pupils with a mean of 18 pupils per class ($SD = 6.71$ pupils), and lessons were taught predominantly in Welsh.

The 12 classes had a total of 16 teachers; 8 teachers shared teaching duties with another teacher, and 8 teachers were full-time teachers. Data from the 8 full-time teachers and 4 of the teachers who shared teaching duties and spent the larger amount of time with their class are reported and analyzed. All teachers were women, with a mean age of 34 years (range 21 to 53 years) and an average of 9 years teaching experience (range 2 to 30 years). Schools were based in both rural and urban areas in the county of Gwynedd in North West Wales.

The percentage of pupils entitled to free school meals ranged from 4% to 27% (mean of 13%) per school; the national average for Wales in 2005/06 (when this study commenced) was 17% (National Statistics, 2006). Although the mean percentage of free school meals was slightly below the national average, the range across the target schools was wide, reflecting the diverse socioeconomic status of school catchment areas. Each participating school, teacher, parent and child was given a small gift for taking part in the study. None of the teachers had previously undertaken any IY training, and none of the schools had any teachers

Table 1

Child, teacher, and school demographic characteristics by condition and across conditions at baseline.

Demographic characteristics	Intervention	Control	All	<i>p</i>
Child				
<i>N</i>	53	54	107	
Age, months: <i>M (SD)</i>	58.23 (7.66)	56.76 (5.73)	57.51 (6.80)	.272 ^a
Gender				
Boys, <i>n (%)</i>	26 (45)	32 (55)	58 (54)	.290 ^b
Girls, <i>n (%)</i>	27 (55)	22 (45)	49 (46)	
Teacher				
<i>N</i>	6	6	12	
Age, years: <i>M (SD)</i>	37.3 (12.69)	32.3 (9.77)	34.83 (11.11)	.463 ^a
Gender, <i>n (%)</i>				
Women	6 (100)	6 (100)	12 (100)	
No. years teaching, <i>M (SD)</i>	7.67 (5.32)	11.00 (10.00)	9.33 (7.83)	.492 ^a
No. schools taught, <i>M (SD)</i>	3.33 (2.16)	1.50 (0.55)	2.42 (1.78)	.094 ^a
Highest position, <i>n (%)</i>				
Class teacher	4 (66)	4 (66)	8 (66)	
Head infant school	1 (17)	1 (17)	2 (17)	
Acting head school	1 (17)	1 (17)	2 (17)	
Class taught, <i>n (%)</i>				
Reception age	3 (50)	2 (33)	5 (42)	.558 ^b
Multi-age	3 (50)	4 (67)	7 (58)	
Classroom				
<i>N</i>	6	6	12	
Class size: <i>M (SD)</i>	18.17 (6.82)	17.83 (7.25)	18.00 (6.71)	.936 ^a
School				
<i>N</i>	5	6	11	
Total no. pupils, <i>M (SD)</i>	107.80 (67.44)	106.00 (65.55)	106.82 (63.00)	.965 ^a
Rurality, <i>n (%)</i>				
Rural	2 (40)	2 (33)	4 (36)	.946 ^b
Semi-rural	1 (20)	1 (17)	2 (18)	
Urban	2 (40)	3 (50)	5 (46)	

Note.

^a Significant differences tested using an independent sample *t*-test.^b Significant differences tested using a chi-square test of independence.

that had been trained in the IY-TCM or child interventions prior to the commencement of the trial. Five of the six intervention teachers attended all five training sessions; the remaining teacher attended four of the five sessions. All teachers were retained for the duration of the study. One school had two classrooms in the study. Both of these were in the intervention condition; therefore, there were no issues regarding cross-contamination between classrooms.

Teachers rated all children in their class using the teacher version of the Strengths and Difficulties Questionnaire (TSDQ; Goodman, 1997). One hundred and seven children (58 boys and 49 girls) with an age range of 3 to 7 years ($M = 57.5$ months, $SD = 6.80$ months) were recruited on the basis of their score on the TSDQ (see screening measure subsection). All included children were of Caucasian ethnicity.

A total of 107 children were rated by teachers at preintervention. Because nine children were recruited from each classroom, it was expected that they would be the three highest, three midpoint, and three lowest scoring children; however the sample as a whole included 80 children (75%) whose scores fell in the low scoring range on the TSDQ and 27 children (25%) whose scores fell within the high scoring range on the TSDQ. A total of 40 boys and 40 girls had scores that fell within the low scoring range, whereas 18 boys and 9 girls had scores in the high scoring range. For data analysis purposes, target children were divided into two groups according to TSDQ guidelines, using the preintervention data. The high-scoring group included 18 children who were rated as above the point of clinical concern ($TSDQ > 15$; see www.sdqinfo.com for more information). There was a mean of 2.33 high-scoring children per class with a range of 0 to 4 children per class.

2.2. Intervention

The IY-TCM training was delivered to teachers during one full-day session each month over the course of 5 months, after initial observations and prior to postintervention observations, which were all carried out during the same school year. The five-session program was used due to the fact that this was the published version of the IY-TCM program available at the time of the study (see Table 2 for a comparison of the five-session and six-session versions of the IY-TCM program). In the case of the two classes that had job-share teachers and were allocated to the intervention condition, both teachers attended the training. Two trained leaders, a certified program mentor (the first author) and a trained program leader, ran the course. The first author had completed the leader certification process that involved submission of a recording of a session of the program and records of teacher responses to each session from two complete programs. To become a certified program mentor, she had also been trained to deliver leader training in the program to future program leaders. The coleader had also delivered the course on several prior

Table 2

Contents of the 5-session (pre 2006) and 6-session (post 2006) versions of the Teacher Classroom Management (TCM) training.

TCM 5-session training	Current TCM 6-session training
<u>Session 1</u> Topic of Morning – Building positive relationships with students Topic of Afternoon – Proactive teacher: Preventive approaches	<u>Session 1</u> Topic of Morning – Building positive relationships with students Topic of Afternoon – Proactive teacher: Preventive approaches
<u>Session 2</u> Topic of Morning – Teacher attention, praise, and encouragement Topic of Afternoon – Teacher coaching, child-directed play, and friendship skills	<u>Session 2</u> Topic of Morning – Teacher attention, praise, and encouragement Topic of Afternoon – Teacher coaching, child-directed play, and friendship skills
<u>Session 3</u> Topic of Morning – Motivating students through incentives Topic of Afternoon – Dialogic reading	<u>Session 3</u> Topic of Morning – Motivating students through incentives Topic of Afternoon – Motivating students continued
<u>Session 4</u> Topic of Morning – Decreasing inappropriate behavior Topic of Afternoon – Decreasing inappropriate behavior continued	<u>Session 4</u> Topic of Morning – Decreasing inappropriate behavior Topic of Afternoon – Decreasing inappropriate behavior continued
<u>Session 5</u> Topic of Morning – Teaching children to be socially competent Topic of Afternoon – Teaching children to be socially competent continued	<u>Session 5</u> Topic of Morning – Decreasing inappropriate behavior continued Topic of Afternoon – Teaching children to be socially competent
N/A	<u>Session 6</u> Topic of Morning – Teaching children to be socially competent continued Topic of Afternoon – The complete teacher: Emotion Regulation and problem solving

occasions. See www.incredibleyears.com for further details on the leader certification process. Fidelity in training was addressed by means of a series of checklists (both leader- and teacher-completed) as identified by the program manual (Webster-Stratton, 2003) to ensure evidence-based implementation. These were subsequently used to guide future training sessions. Teachers also completed a satisfaction questionnaire at the end of the course (see Measures section).

At each session, teacher participants identified key classroom management principles from discussions based on their own experience, observed and discussed video footage of classroom situations, participated in role-play, and undertook behavior planning for specific pupils. At the end of each session, they were given classroom assignments (including implementation of a behavior plan for a target pupil) to complete during the subsequent month. Teachers were instructed to practice the program principles during the month following each session and to report back on their experiences and behavior plans at the start of the following session. No ongoing coaching by program leaders was provided during the month following each session; however, all teachers had the opportunity to contact the program leaders if necessary.

The program aimed to teach the teacher how to develop a positive relationship with children through play, use specific labeled praise for appropriate behavior (e.g., “Well done for putting your books away when I asked you”), provide incentives for target behaviors that were difficult for the child, give clear commands, ignore minor inappropriate behavior and encourage other pupils to do likewise, use positive discipline strategies, and promote emotion regulation and problem solving with pupils. Investment in developing a positive relationship with children is the foundation on which all of the other program principles are built (Webster-Stratton & Reid, 2002). The methods used during the five-session TCM program are the same as those now used in the six-session TCM program. These methods include videotape modeling through the use of vignettes, practicing and rehearsing through role-plays and discussions, developing individual behavior plans, and giving homework assignments at the end of every session so that teachers can practice new skills in their own classrooms between sessions.

2.3. Measures

Child behaviors were assessed by means of a questionnaire. In addition, child behaviors and teacher behaviors were assessed on two separate occasions, preintervention and postintervention (see Procedure for further detail). Fidelity was assessed through the use of checklists and an end of course satisfaction questionnaire.

2.3.1. Teacher version of the Strengths and Difficulties Questionnaire (TSDQ; Goodman, 1997)

The TSDQ is an established 25-item rating scale that screens for hyperactivity problems, peer problems, emotional problems, and conduct problems, as well as prosociality. The four problem subscales' scores are combined to produce a Total Difficulties score. The TSDQ Total Difficulties score displays moderate internal consistency ($\alpha = .82$), good test–retest reliability ($r = .84$), and concurrent relations with the Child Behavior Checklist total score ($r = .76$; Stone, Otten, Engels, Vermulst, & Janssens, 2010). The Total Difficulties score has borderline and clinical level cut-off scores: 0–11 low, 12–15 borderline, and 16–40 clinical. For this study, children in the low score group had TSDQ Total Difficulties scores of 0–11, while those in the high score group had TSDQ Total Difficulties scores above 12.

2.3.2. The Teacher–Pupil Observation Tool (TPOT; Martin et al., 2010)

The TPOT (Martin et al., 2010) is a 75-item measure of the frequency of teacher and child behaviors of which half of the items are devoted to child interactions or responses (with or to the teacher or other pupils in the class) and the other half of the items

are devoted to teacher interactions (with the target child or with other pupils in the class). This measure was designed specifically to capture the core skills taught during the IY-TCM training. Based on previous analysis of the TPOT, the 75 items were reduced to form nine composite categories: teacher positives; teacher negatives; teacher commands; compliance; noncompliance; negative behaviors towards the teacher; prosocial behavior; deviance, and off-task behavior (see Martin et al., 2010). Additionally, children's percentage compliance to teacher commands was examined using the following formula – child compliance divided by teacher commands, multiplied by 100. The TPOT displays good inter-rater agreement, with a mean intraclass correlation of .78 across all 75 categories; small internal consistency, $\alpha = .49$; and moderate concurrent validity with SDQ categories, $u = .41$ (u = effect size for Mann–Whitney U; above .1 is a *small* effect size, .3 is *moderate*, and anything above .5 is *large*). Scores on the TPOT have been shown to exhibit good discriminant validity, with high scores on negative categories strongly associated with increased behavioral, social, and emotional problems, and high scores on positive categories strongly associated with high prosociality (Martin et al., 2010).

2.3.3. Fidelity measures

This study included three fidelity measures. All three fidelity measures do not have psychometric support.

2.3.3.1. Teacher Satisfaction Questionnaire (Webster-Stratton & Reid, 2002). The Teacher Satisfaction Questionnaire includes five items. The first four items cover four areas: (1) usefulness of the program, (2) confidence in using the ideas, (3) difficulty/ease of putting the program into practice, and (4) use of strategies for improving home–school links. A final item asks teachers to describe the two most useful things they had taken from the program. Items are rated on a seven-point scale with one being a negative response and seven being a positive response.

2.3.3.2. Incredible years teacher workshop evaluations (Webster-Stratton & Reid, 2002). At the end of every workshop session, participating teachers completed the IY Workshop Evaluation form. This form has four items regarding the helpfulness of the content of the session, the videotape examples shown during the session, the group leaders' teaching, and the group discussions. Responses are rated on a four point scale comprising of *not helpful*, *neutral*, *helpful*, and *very helpful*.

2.3.3.3. Teacher workshop checklist (Webster-Stratton & Reid, 2002). At the end of every workshop session, group leaders completed the Teacher Workshop Checklist for that particular session. This form is used to check whether group leaders are adhering to program fidelity such as showing the necessary video vignettes, engaging in recommended practice exercises and brainstorming, and using the key learning principles. Items include circling the specific vignettes used during the session (numbers ranged from 1 to 57 depending on session) and Yes/No response items regarding the different methods used during the session. Examples include “Write agenda on the board”, “Brainstorm ideas for building relationships with students,” and “Role play descriptive commenting.” Numbers of Yes/No items ranged from 1 to 9 depending on session.

2.4. Procedure

2.4.1. School-based implementation

A Gwynedd Education Service official recruited the schools. Gwynedd was in the process of implementing the IY-TCM program in all 102 of its primary schools, and at the start of the study, it had been implemented in 16 schools. Incentives to participate in the research were the opportunity to receive the training either earlier or at the same time as other schools in the same catchment area and funding for substitute teachers to cover teacher attendance at the IY-TCM course. A total of 13 schools gave permission to be contacted by the research team. Of these, two declined to take part; one was in a catchment area that had their training moved forward by the local education authority, while the other had staffing problems.

The study was conducted within one whole school year. Between September and October, teachers completed the TSDQ for each pupil in their class. Nine children were recruited from each classroom, with the exception of one classroom from which only eight children were recruited. This exception was a multiage classroom that was recruited later than the others, and consequently, only eight parents had consented to their child's participation by preintervention assessments. An independent researcher scored the TSDQs and contacted the schools with the age and initials of nine identified children (the three highest, the three lowest, and three mid-range scoring children) so as to ensure that the coding team would remain blind to child scores. It was decided not to include all children because limited resources meant that it was not feasible to complete classroom observations on all children in the classrooms. They were recruited to provide a representative sample to enable assessment of the impact of the TCM program on children with different levels of behavioral difficulties. Teachers subsequently contacted each potential child's parents to explain the research and to obtain consent to the classroom observation. Parents were told that their teacher was participating in a research trial and that, in order to assess the outcome, the teacher would be observed interacting with children in the classroom. If parental consent to observation of a specific child was not given, the parent of the child with the next nearest score was approached. Eleven parents (9%) did not consent for their child to be individually observed, and seven (64%) of these parents were the parents of children with the highest TSDQ Total Difficulties scores in their respective classrooms. Of the remaining four parents who did not consent to individual observations, two (18%) were the parents of children with the lowest TSDQ Total Difficulties scores in their classrooms, and two (18%) were the parents of the children with TSDQ Total Difficulties scores in the middle range.

2.4.2. Assessment

In November and December, preintervention classroom observations were conducted. IY-TCM training then took place during January to May. Postintervention classroom observations were conducted in June and July. It was decided that postintervention measures would be completed one month after the end of the IY-TCM training due to the ongoing nature of the training and within the same school year to ensure that the same teacher was observed with the same children. All the teachers were satisfied with the IY-TCM training provided with all teachers rating the different components as either *helpful* or *very helpful*. No variations in training across sites were reported by group leaders.

Three postgraduate students (all with a master's degree) with previous experience of similar observation measures conducted classroom observations. Their initial training comprised studying the TPOT manual, observing and coding video footage, and comparing frequency counts until they had reached reliability levels of 70% agreement with the primary coder (i.e., the second author). Top-up sessions of 30 to 60 min were held weekly in order to prevent coder drift and to discuss any questions arising from recent classroom visits. All TPOT behaviors were coded by means of frequency counts representing each occurrence.

The observations were conducted during structured lessons such as math/numbers or reading/writing. All but one teacher delivered these structured lessons in the morning, the remaining teacher was observed in the afternoon. Each target child was observed for 15 min, a time period that was determined as adequate following a previous study (see Martin et al., 2010). Observers also simultaneously coded teacher interactions with the class as a whole and class responses to the teacher and behavior generally. All three observers observed in all of the classrooms apart from three classrooms where only two observers were present. Observers remained as unobtrusive as possible, keeping interactions to a minimum so as not to affect the behavior of children in the classroom. Observers maintained reasonably high inter-rater reliability (86%) by coding videotapes of classroom interactions between actual observations. Observations of the same child in the classroom were double coded (by two or more observers) for 25% of observations (86% reliability). Observations at preintervention were undertaken prior to randomization, and observers were blind to both children's TSDQ scores and teachers' group allocation (intervention or control) at postintervention.

2.5. Randomization

This study adopted a single-blind stratified RCT design. An independent researcher paired classrooms according to school size, classroom size, and locality (town/rural) after preintervention measures had been collected. Schools were subsequently randomly allocated to the intervention or control condition using a random number generator; the paired school was automatically allocated to the control group in each case. Control teachers were offered the IY-TCM training in the following academic year. No in-service or professional development opportunities were provided to the control group schools during the study.

2.6. Analysis strategy

All TPOT category variables were examined for normality. Graphical analysis of residuals showed the assumption of normality and equal variability approximately held for all TPOT variables. Before the main analyses were conducted, baseline differences between conditions were examined using independent sample *t*-tests. All 16 teachers (including job-share teachers) were observed, but data analysis was based on the 12 teachers (4 intervention and 4 control full-time teachers and 2 job-share intervention and 2 control teachers who taught the greater part of the school week). This restricted data analysis ensured that data from each child was entered only once in the analysis of outcomes for the target children.

To examine the effect of intervention on teacher and child behavior, hierarchical linear modeling (HLM) analyses were conducted using SPSS 17.0 to account for the clustered nature of the data. A two-level model was examined: children at Level 1 and classrooms at Level 2. The dependent variables were the postintervention scores for the TPOT categories. Covariates included preintervention scores at Level 2 and gender, preintervention age, and preintervention TSDQ Total Difficulties score at Level 1. Condition-by-TSDQ Total Difficulties score was included in the Level 1 model to examine whether intervention effects differed for children with high versus low problem behaviors. All continuous variables were centered (within the classroom cluster). The equation for the Level 1 model was specified as follows:

$$Y_{ij} = b_{0j} + b_{1j}(\text{TSDQ TOTAL})_{ij} + b_{2j}(\text{PRE SCORE})_{ij} + b_{3j}(\text{AGE})_{ij} + b_{4j}(\text{SEX})_{ij} + e_{ij}.$$

The Level 2 model was specified as follows:

$$\begin{aligned} b_{0j} &= g_{00} + g_{01}\text{CONDITION}_j + g_{02}\text{PRE SCORE}_j + m_{0j} \\ b_{1j} &= g_{10} + g_{11}\text{CONDITION}_j \\ b_{2j} &= g_{20} \\ b_{3j} &= g_{30} \\ b_{4j} &= g_{40}. \end{aligned}$$

Y_{ij} represents the postintervention TPOT score for student i in classroom j , β_{0j} represents the intercept, SEX was a dichotomous variable where children were coded with a zero for a male and one for a female, AGE was a continuous variable measured in months, TSDQ TOTAL was a dichotomous variable where children were coded as 0 if they had a TSDQ Total Difficulties score of 12

Table 3

Hierarchical linear modeling results for classroom outcomes (Level 2) with descriptive statistics by condition and effect sizes.

Outcome ^a	Estimate	Std. Error	ICC	<i>p</i>	Intervention (<i>n</i> = 6)		Control (<i>n</i> = 6)		<i>d</i>
					Pre	Post	Pre	Post	
					<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	
Teacher positive to class	9.23	10.41	.39	.392	83.09 (17.29)	97.32 (15.38)	89.15 (24.51)	88.02 (23.24)	0.17
Teacher negative to class	−0.48	3.44	.61	.891	10.11 (7.51)	8.22 (6.04)	7.37 (5.31)	8.71 (7.02)	0.03
Class compliance	3.69	6.19	.29	.563	40.84 (14.47)	53.57 (13.74)	47.83 (17.22)	49.71 (9.66)	0.12
Class noncompliance	−0.44	0.29	.01	.151	1.16 (1.26)	.66 (.56)	.69 (.37)	1.09 (.52)	0.30
Class negatives to teacher	−1.25	1.10	.23	.277	1.93 (2.07)	1.51 (1.57)	3.74 (1.89)	2.79 (2.53)	0.22
Class prosocial behavior	−2.78	4.36	.19	.536	25.52 (7.19)	23.89 (5.54)	30.24 (9.26)	26.78 (10.38)	0.12
Class deviant behavior	−0.65	1.71	.11	.713	4.45 (2.89)	4.24 (3.28)	7.43 (4.33)	4.90 (3.29)	0.07
Class off-task behavior	−0.49	0.18	.06	.016*	.53 (.36)	.25 (.28)	.33 (.29)	.77 (.42)	0.53

Note. ICC = intraclass correlation coefficient; TSDQ = Teacher Strengths and Difficulties Questionnaire.

^a All outcome variables were scored from the Teacher–Pupil Observation Tool.

* *p* < .05.

or above and 1 if they had a TSDQ Total Difficulties score below 12. CONDITION was a dichotomous variable where 0 represented the control group and 1 represented the intervention group.

Analyses were conducted using full maximum estimation likelihood to account for missing data under an assumption that data were missing at random. Only one child in the control group was lost to follow-up (1%). Complete data were available for the TSDQ TOTAL and PRE SCORES covariates; however, there were missing data for the GENDER (4%) and AGE (4%) covariates. Intraclass correlation coefficients (ICCs) were calculated to estimate the proportion of variance in outcomes due to cluster effects.

Table 4

Hierarchical linear modeling results for index child outcomes (Level 1) with main effects and condition × TSDQ total score interaction.

Outcome	Estimate	Std. Error	ICC	<i>p</i>	<i>d</i>
Teacher positive to index child			.31		
Condition	4.37	5.19		.404	0.03
Cond * TSDQ TOTAL	−7.50	4.81		.122	
Teacher negative to index child			.53		
Condition	−1.17	.057		.044*	0.36
Cond * TSDQ TOTAL	0.97	0.61		.116	
Child compliance			.15		
Condition	−6.37	3.10		.045*	0.37
Cond * TSDQ TOTAL	4.55	3.25		.165	
Child noncompliance			.00		
Condition	−0.59	0.37		.115	0.36
Cond * TSDQ TOTAL	0.38	0.43		.374	
Child negative to teacher			.28		
Condition	−1.28	0.47		.011*	0.42
Cond * TSDQ TOTAL	0.96	0.42		.024*	
Child prosocial behavior			.20		
Condition	−2.97	3.17		.356	0.23
Cond * TSDQ TOTAL	.39	3.06		.899	
Child deviant behavior			.00		
Condition	−1.57	0.84		.066	0.38
Cond * TSDQ TOTAL	1.23	0.98		.214	
Child off-task behavior			.14		
Condition	−1.45	0.46		.003**	0.48
Cond * TSDQ TOTAL	1.37	0.48		.005**	
Total number commands			.16		
Condition	−10.22	4.00		.014*	0.48
Cond * TSDQ TOTAL	6.68	4.06		.104	

Note. ICC = intraclass correlation coefficient; TSDQ = Teacher Strengths and Difficulties Questionnaire.

* *p* < .05.

** *p* < .01.

Table 5
Descriptive statistics for outcomes related to index children by condition, TSDQ total score and overall.

Outcome ^a	Intervention Group						Control Group					
	All (N = 53)		Low TSDQ (n = 40)		High TSDQ (n = 13)		All (N = 54)		Low TSDQ (n = 40)		High TSDQ (n = 14)	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)
Teacher positives to index child	17.40 (14.86)	13.28 (13.13)	17.47 (16.02)	9.55 (9.34)	17.15 (11.08)	24.77 (16.57)	13.06 (12.90)	14.45 (13.22)	12.92 (11.37)	13.13 (11.86)	13.43 (17.04)	18.14 (16.35)
Teacher negatives to index child	1.34 (3.36)	0.60 (1.67)	1.13 (3.49)	0.50 (1.77)	2.00 (2.94)	0.92 (1.32)	0.48 (0.84)	1.02 (1.66)	0.38 (0.67)	0.77 (1.16)	0.79 (1.19)	1.71 (2.52)
Child compliance	8.85 (8.74)	7.62 (7.35)	8.43 (9.21)	6.32 (6.73)	10.15 (7.29)	11.62 (7.99)	8.28 (7.20)	9.60 (8.77)	8.28 (6.92)	7.6 (6.81)	8.29 (8.20)	14.93 (11.41)
Child noncompliance	0.62 (2.51)	0.13 (0.39)	0.63 (2.84)	0.10 (0.30)	0.62 (0.96)	0.23 (0.60)	0.09 (0.29)	0.42 (1.28)	0.08 (0.27)	0.28 (1.30)	0.14 (0.36)	0.79 (1.19)
Child negatives to teacher	0.26 (0.84)	0.19 (0.52)	0.05 (0.32)	0.18 (0.50)	0.92 (1.44)	0.23 (0.60)	0.43 (0.90)	0.62 (1.35)	0.28 (0.60)	0.44 (1.17)	0.86 (1.41)	1.14 (1.70)
Child prosocial behavior	9.26 (7.23)	7.21 (5.13)	9.48 (7.49)	7.10 (4.54)	8.62 (6.59)	7.54 (6.85)	10.09 (7.91)	9.94 (9.01)	10.48 (8.41)	10.56 (9.83)	9.00 (6.41)	8.21 (6.17)
Child deviance	0.98 (2.12)	0.58 (1.03)	0.65 (1.78)	0.60 (1.03)	2.00 (2.77)	0.54 (1.05)	1.70 (3.87)	1.21 (2.81)	1.10 (1.71)	1.05 (2.81)	3.43 (6.93)	1.64 (2.87)
Child off-task behavior	0.94 (2.42)	0.45 (0.93)	0.40 (0.93)	0.45 (0.93)	2.62 (4.31)	0.46 (0.97)	0.94 (1.71)	0.85 (1.35)	0.70 (1.47)	0.59 (1.21)	1.64 (2.17)	1.57 (1.51)
Total commands	13.06 (12.07)	9.36 (8.70)	12.30 (12.59)	7.43 (7.67)	15.38 (10.43)	15.31 (9.26)	11.70 (10.79)	13.62 (12.24)	11.73 (10.73)	10.59 (9.04)	11.64 (11.36)	22.07 (16.01)

Note. TSDQ = Teacher Strengths and Difficulties Questionnaire.

^a All outcome variables were scored from the Teacher–Pupil Observation Tool.

Standardized mean differences were computed for the effect size using standard deviations generated in the HLM analyses. An alpha level of .05 was used for all tests of significance.

3. Results

3.1. Preintervention analyses

There were no significant differences in child, teacher, or school demographics between control and intervention groups preintervention (see Table 1). There were no significant differences in the number of children in the low and high TSDQ groups between the intervention and control conditions preintervention, $\chi^2(1) = 0.03, p = .868$. All subsequent analyses controlled for the effects of preintervention scores centered within classroom cluster.

3.2. Classroom-level findings

Table 3 presents the results of the intervention on teacher and classroom behavior. There was a significant main effect of condition for classroom off-task behavior. Classrooms in the intervention condition showed significant reductions in off-task behavior ($d = 0.53$). All other TPOT categories showed no significant differences between conditions at postintervention.

3.3. Child-level findings

Table 4 presents the results of the intervention on teacher and target children's behavior. Means and standard deviations are displayed in Table 5. There was a significant main effect of condition on teacher negatives; teachers in the intervention group showed reduced levels of negatives at postintervention ($d = 0.36$). There was also a significant main effect of condition for both child negatives to the teacher and child off-task behavior; children in the intervention group showing reductions in both child negatives to teacher ($d = 0.42$) and child off-task behavior ($d = 0.48$).

Additionally, there was a significant main effect of condition on child compliance with children in the intervention group showing a significant reduction in the number of compliances to commands ($d = 0.37$). Because one of the aims of the IY-TCM intervention is to reduce the number of commands given by teachers and, in turn, increase child compliance, further analyses were conducted to explore these data (see the bottom rows in Tables 4 and 5). Analyses showed a significant reduction in teacher commands to children in the intervention group ($d = 0.48$), while teachers in the control group showed an increase in the total number of commands given. Furthermore, percentage compliance to teacher commands (Table 6) showed that the percentage of compliance to commands by children in the intervention had increased from 68% to 81%, whereas percentage compliance for children in the control group showed no change.

Two condition-by-TSDQ Total Difficulties interaction effects were found to be significant (see Table 4). Graphical representations of the results are displayed in Figs. 1 and 2. Fig. 1 shows the condition-by-TSDQ Total Difficulties score interaction for child negatives to teacher. Children in the control group, irrespective of TSDQ Total Difficulties score, showed an increase in negatives to the teacher from preintervention to postintervention. Children in the intervention group, on the other hand, showed a different pattern; children with low scores displayed similar patterns to children in the control group. However, children with high scores showed a decrease in negatives towards the teacher. Fig. 2 shows the condition-by-TSDQ Total Difficulties score interaction for child off-task behavior. Children in the control group, irrespective of TSDQ Total Difficulties score, showed no change in off-task behavior from preintervention to postintervention. Children in the intervention group with low scores also did not show any change in their off-task behavior. However, children in the intervention group with high scores showed a decrease in their off-task behavior.

4. Discussion

The purpose of this RCT was to evaluate the effectiveness of the IY-TCM program as a stand-alone universal program for primary school teachers. Four hypotheses were tested: (a) teachers in the intervention group would display more positive and less negative behaviors towards the whole class, (b) teachers in the intervention condition would show more positive and less negative behaviors towards children with high TSDQ Total Difficulties scores, (c) children in intervention classrooms would show more positive and less negative behaviors than those in control classrooms, and (d) children with high TSDQ Total Difficulties

Table 6

Percentage compliance to total number of teacher commands by condition.

Compliance variables	Intervention ($n = 53$)		Control ($n = 54$)	
	Pre $M (SD)$	Post $M (SD)$	Pre $M (SD)$	Post $M (SD)$
Teacher commands	13.06 (12.07)	9.36 (8.70)	11.70 (10.79)	13.62 (12.24)
Child compliance	8.85 (8.74)	7.62 (7.35)	8.28 (7.20)	9.60 (8.77)
Percentage compliance (%)	68	81	71	71

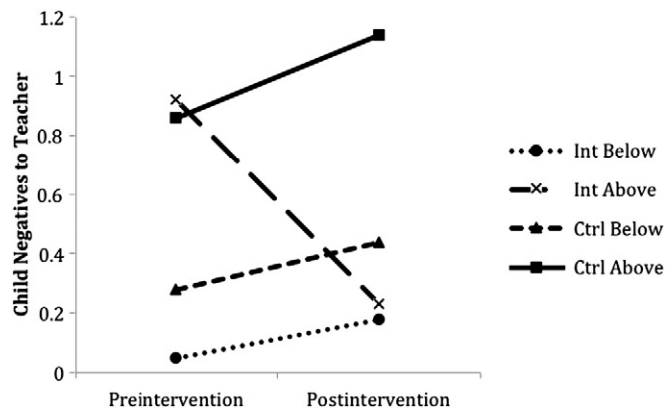


Fig. 1. Condition * TSDQ Total Difficulties score interaction for Child Negatives to Teacher. TSDQ = Teacher Strengths and Difficulties Questionnaire; Int Below = children in the intervention group scoring within the low range on the TSDQ Total Difficulties; Int Above = children in the intervention group scoring within the high range on the TSDQ Total Difficulties; Ctrl Below = children in the control group scoring within the low range on the TSDQ Total Difficulties; Ctrl Above = children in the control group scoring within the high range on the TSDQ Total Difficulties.

scores in intervention classrooms would show greater behavioral changes compared to children with high TSDQ Total Difficulties scores in control classrooms. Data analyses using HLM showed no significant differences in teachers' behavior towards the whole class yet significant reductions in negative behavior of teachers in the intervention group towards target children, regardless of risk status. Children in intervention classrooms showed reductions in off-task behavior, and high-risk children in the intervention condition showed significant reductions in negatives towards the teacher and off-task behavior.

4.1. Effects on teacher behavior

The first hypothesis was not supported because there was no significant change in teacher behavior towards the whole class. There are several possible reasons for this finding. It is possible that the teachers gave more attention to target children because they were aware that the observers were observing the target children. Teachers would therefore be more likely to have positive interactions with the target children as opposed to the other children in the classroom. Another possibility concerns measurement issues. It could be that the TPOT is not sensitive enough to capture changes in teacher behavior at the classroom level. Further research would need to be conducted to examine the TPOT's sensitivity to changes at the classroom level.

The second hypothesis was partially supported because there was a significant reduction in teacher negative behavior towards target children in the intervention group; however, it was not related to child behavior problem scores. Nevertheless, similar findings have been reported in previous research on the IY-TCM program (e.g., Baker-Henningham et al., 2009; Webster-Stratton et al., 2004; Webster-Stratton et al., 2008). These findings also support research conducted with other classroom management programs. One example is Leflot, Van Lier, Onghena, and Colpin (2010), who found that teachers who delivered the Good Behavior Game used fewer negative remarks when managing children's disruptive behavior compared to teachers in the control

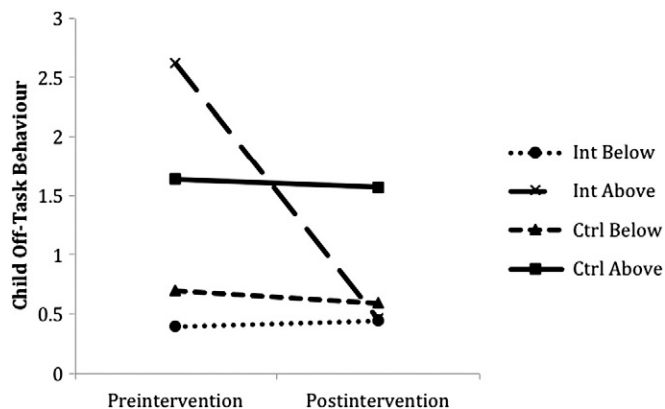


Fig. 2. Condition * TSDQ Total Difficulties score interaction for Child Off-Task Behavior. TSDQ = Teacher Strengths and Difficulties Questionnaire; Int Below = children in the intervention group scoring within the low range on the TSDQ Total Difficulties; Int Above = children in the intervention group scoring within the high range on the TSDQ Total Difficulties; Ctrl Below = children in the control group scoring within the low range on the TSDQ Total Difficulties; Ctrl Above = children in the control group scoring within the high range on the TSDQ Total Difficulties.

group. Because the intervention did lead to a reduction in teacher negatives to target children, it emphasizes the effectiveness of the IY-TCM program as a universal program for all classrooms and all children regardless of behavior problem risk.

4.2. Effects on classroom behavior

There was a significant reduction in the off-task behavior of children in intervention classrooms, which partially supports the third hypothesis. Similar findings have been reported in previous IY-TCM trials, such as Webster-Stratton et al. (2008) who found improvements in the levels of disengagement of children in intervention classrooms. Previous research has shown that poor classroom management skills are associated with child negative behaviors such as off-task behavior, which in turn predicts poor academic achievement (Webster-Stratton et al., 2008). The IY-TCM program emphasizes the importance of proactive teaching practices to prevent disruptive and off-task behaviors (Webster-Stratton & Reid, 2002). Teachers are encouraged to establish positive classroom rules, to have predictable school routines, to use clear commands to ensure that children know what to expect and that they stay engaged in the task at hand, and to ensure that commands are really necessary before giving them. Targeting poor classroom practices may be one way of improving children's academic achievements by reducing child negative behaviors such as off-task behavior as shown in the current study.

4.3. Effects on target child behavior

We found a significant main effect of condition on target child compliance to commands, with children in the intervention showing lower levels of compliance to commands at postintervention. However, further analyses revealed that teachers in the intervention group gave significantly fewer commands to target children than teachers in the control group. Percentage compliance to teacher commands also increased for children in the intervention group, while there was no change for children in the control group. One of the aims of the IY-TCM intervention is to reduce the number of commands given by teachers and, in turn, increase child compliance (Webster-Stratton & Reid, 2002). This aim is supported by the results of this study. Other studies have examined whether the IY-TCM intervention leads to a reduction in the number of commands by teachers; however, they found no significant difference postintervention (Baker-Henningham et al., 2009; Hutchings et al., 2007).

The current study is the first to demonstrate that the IY-TCM program can lead to reductions in the total number of commands given to target children and in turn lead to an increase in the rate of compliance. Giving only a few clear, specific commands means that children know exactly what is expected of them; such commands can reduce the occurrence of behavior problems in the classroom. Giving clear, specific commands is one of the strategies used for proactive teaching (Webster-Stratton & Reid, 2002) taught in the program. Carlson et al. (2011) found increases in proactive strategies six months after completing the IY-TCM training, which incorporates using fewer, clear, specific commands as opposed to many nonspecific commands. Raver et al. (2008) also found statistically significant increases in teachers' proactive behavior management practices following the IY-TCM training with added consultation support for teachers. Proactive teaching strategies have been linked with improved child self-regulation (Webster-Stratton & Taylor, 2001) and positive classroom climate (Raver et al., 2008).

The intervention also had a positive main effect on child negatives towards the teacher and child off-task behavior, showing significant reductions in both. Other studies investigating the IY-TCM program report similar findings. All three RCTs by Webster-Stratton and colleagues found significant observed reductions in child conduct problems in school, which included negative behaviors towards teachers (Webster-Stratton et al., 2001a; Webster-Stratton et al., 2001b; Webster-Stratton et al., 2004, 2008). Additionally, Webster-Stratton et al. (2008) found a significant reduction in off-task behavior. Shernoff and Kratochwill (2007), however, found no reduction in disruptive behavior using an observation tool that included negative social engagement as a category. The finding of a reduction in both negatives towards the teacher and off-task behavior may be due to a reduction in teacher negatives to the target children leading to increased motivation in the children.

Additionally, interaction effects for negative behavior towards the teacher and off-task behavior were found for target children. Both variables showed similar behavioral patterns with children in the control group, regardless of TSDQ Total Difficulties score. In a similar manner, children in the intervention group with low TSDQ Total Difficulties scores showed no change at postintervention. Children in the intervention group with high TSDQ Total Difficulties scores, on the other hand, showed a significant reduction in both negative behavior towards the teacher and off-task behavior. These findings support the fourth hypothesis. Previous research has found similar results. In particular, Webster-Stratton et al. (2004) found that children at high-risk of conduct problems showed significant reductions in negative classroom behavior after their teachers engaged in the IY-TCM program. Again, these results suggest that behavioral changes in children are possible without delivering an intervention directly with the children.

Interestingly, children in the intervention group with low TSDQ Total Difficulties scores showed no change at postintervention. They seem to show deterioration in behavior with a slight increase in negatives to the teacher and a decrease in compliance. This finding could be due to a decrease in interaction between these children and the teacher, which is supported by a decrease in teacher positive and negative behavior towards low-risk children in the intervention group. Teachers may be spending more time with the high scoring children over the course of the school year in order to manage their classroom behavior. A similar pattern can be seen in the control group in that there was more teacher interaction with the children with high TSDQ Total Difficulties scores than the children with low TSDQ Total Difficulties scores at postintervention. In a similar manner, low scoring children in the control group also showed deterioration in their behavior over the course of the school year. For this study, teachers were

aware of the identity of the target children being observed in the classroom. This awareness could have affected the way the teacher interacted with the children in the classroom.

Collectively, these results strengthen the growing evidence for the effectiveness of the IY-TCM program (e.g., Webster-Stratton et al., 2001a; Webster-Stratton et al., 2001b; Webster-Stratton et al., 2004, 2008; Baker-Henningham et al., 2012; Raver et al., 2008; Shernoff & Kratochwill, 2007) particularly because this study is the first to investigate the effectiveness of the IY-TCM program as a stand-alone, nonadapted intervention using an RCT design and independent, blind classroom observations. Combined, these studies show the universality of the program as an effective teacher classroom management program. Significant results have been found with different cultures, including Jamaica (Baker-Henningham et al., 2009, 2012), Wales (Hutchings et al., 2007), and the United States (Raver et al., 2008, Shernoff & Kratochwill, 2007, Webster-Stratton et al., 2001a; Webster-Stratton et al., 2001b; Webster-Stratton et al., 2004, 2008; Williford & Shelton, 2008) as well as with different child behavior problems (e.g., externalizing problems in Webster-Stratton et al., 2004, 2008 and internalizing problems in Herman et al., 2011).

Even though the core components of the IY-TCM are clearly set out in the leader manual, the delivery of the program itself is guided by a set of principles. Having these principles means that the program is flexible and easy to adapt to the needs of the group (e.g., cultural adaptations, variations in teacher skill levels, variations in child development; Reinke, Stormont, Webster-Stratton, Newcomer, & Herman, 2012). A great deal of research has investigated mechanisms of change in the classroom, and while it is clear that negative child behavior in the classroom influences teacher responses (Rimm-Kaufman et al., 2002), this study demonstrated that teacher behavior was a key mechanism for change in the classroom, effectively changing child behavior for the better without the need of a direct intervention with the children. The ability of the intervention to reduce both teacher and child negative behavior appears to be its most important contribution as this enables all children to benefit from the learning environment of the classroom without the added distraction of disruptive behaviors or negative teacher–child interactions. This further improves the classroom learning environment for all children in the class and may help to account for the wider benefits seen.

4.4. Limitations

Although the results were promising there were a number of limitations. First, the sample was small comprising only 12 classrooms and 107 index children. Second, the child age range was narrow (three to seven years) due to the Gwynedd Education Service plan of initially implementing the TCM program with teachers in reception classrooms (four to five years of age) and then subsequently with teachers of older children. Third, the parents of some of the highest scoring children ($n = 7$) in the study did not consent to their children being individually observed. However, those children were included in classroom measures. Although their individual behaviors were not noted as their own, effect of their behaviors on the teacher and the classroom and their interactions with the teacher and target children were recorded and represented in general classroom behavior measures. Fourth, there was no objective measure of treatment integrity as fidelity to the program was only investigated through leader- and teacher-completed checklists, which do not have psychometric support. Fifth, the TSDQ Total Difficulties variable was dichotomized for the analyses, which can reduce power within the analyses. The decision behind dichotomizing this variable was because two of the hypotheses specifically concerned the effects of the intervention on the highest scoring children. Sixth, reliability statistics for the individual composite categories on the TPOT were not available. These issues suggest that caution should be taken in interpreting the results.

4.5. Implications

Teachers are reporting increasing levels of behavior problems in the classroom (Hutchings et al., 2011), and children at highest risk of developing conduct problems are more likely to be taught by teachers who are ill prepared to deal with disruptive classroom behavior (Webster-Stratton et al., 2008). In light of this, improving the classroom management skills of teachers should be a priority for policy makers. This RCT, alongside the previous research, shows that the IY-TCM is an effective way of improving teacher classroom management skills, which in turn reduce child negative behavior in the classroom. National surveys indicate that many schools do not use evidence-based prevention programs (Gottfredson & Gottfredson, 2002). This failure to use such programs could be due to many reasons. For example, professionals in schools may not be aware of available evidence-based programs, or they may lack expertise in identifying and choosing effective programs from among alternatives. Literature reviews have indicated a wide gap between research and practice in school-based prevention methods (Weisz, Sandler, Durlak, & Anton, 2005). The IY-TCM program is an effective evidence-based program that previous research has shown to be acceptable to teachers (Hutchings et al., 2007; Shernoff & Kratochwill) and adaptable to different cultures (Baker-Henningham et al., 2009; Baker-Henningham et al., 2012).

These results added strength to the decision of Gwynedd Education Authority to roll out this program across the county, and since the completion of this study, at least one teacher from all 102 primary schools in the county of Gwynedd in North West Wales has been trained in the IY-TCM program and several schools have been commended for their use of the program by Government Inspectors. Currently evaluations are ongoing in many different countries including England, Norway, Portugal, Ireland, and New Zealand (Hutchings, 2012). The IY-TCM program has also been accredited as a module on the postqualification M.Ed. program at Bangor University enabling access to the training for teachers from across the United Kingdom (Hutchings et al., 2007).

4.6. Future directions

This study was a small-scale RCT examining the effectiveness of the IY-TCM program. Some positive changes were observed in both teacher and child behavior. With a larger sample, it would be possible to explore mediators and moderators of these changes.

The intervention produced some positive changes in child behavior in the classroom; however, it is not known whether these changes extended to other environments, such as the home, where risk factors for conduct problems could persist. Nevertheless, changes in behavior in school do provide some protection for high-risk children even when it is not possible to deliver interventions that target children's out of school circumstances (Webster-Stratton, 1999). Future research should investigate the extent to which changes in child behavior problems within the classroom generalize to the home situation.

There was significantly more postintervention on-task behavior in intervention classrooms, which should predict academic improvements over time. Impact over time of both observed and teacher-reported child behavior could be assessed as could child academic grades to see whether there are identifiable long-term post-TCM training improvements. Additionally, utilizing other rating scales targeting inattention and hyperactivity, such as the Conners rating scales (Conners, Sitarenios, Parker, & Epstein, 1998) and the Kendall Self-Control Rating Scale (Kendall & Wilcox, 1979), would enable assessment of whether the TCM program is an effective in addressing these difficulties. Such effectiveness was established in a study of IY parent training with children at risk of developing ADHD (Jones, Daley, Hutchings, Bywater, & Eames, 2008). Measuring changes in academic performance will also be important within future studies.

Due to funding constraints, there was no long-term follow-up in this study and therefore no evidence as to whether positive changes for the IY-TCM group teachers were maintained with subsequent cohorts of children or whether the changes in child behavior were maintained as they progressed through school. As research highlights the importance of early intervention (Gardner, 2008), exploring the long-term effects of the intervention in a longitudinal study across a wider age range would address generalizability of the intervention across school years. Such research could explore the full age range of the program (3 to 10 years of age) and track child development throughout the school years and across key transition that can pose difficulties for children.

Acknowledgment

This study was funded by an Economic and Social Research Council (ESRC) CASE Award, and co-funded by Gwynedd Education Authority.

References

- Arnold, D. H., Ortiz, C., Curry, J. C., Stowe, R. M., Goldstein, N. E., Fisher, P. H., et al. (1999). Promoting academic success and preventing disruptive behavior disorders through community partnership. *Journal of Community Psychology*, 27, 589–598. [http://dx.doi.org/10.1002/\(SICI\)1520-6629\(199909\)27:5<589::AID-JCOP6-3.0.CO;2-Y](http://dx.doi.org/10.1002/(SICI)1520-6629(199909)27:5<589::AID-JCOP6-3.0.CO;2-Y).
- Baker-Henningham, H., Scott, S., Jones, K., & Walker, S. (2012). Reducing child conduct problems and promoting social skills in a middle-income country: Cluster randomized controlled trial. *The British Journal of Psychiatry*, 201, 101–108. <http://dx.doi.org/10.1192/bjp.bp.111.096834>.
- Baker-Henningham, H., Walker, S., Powell, C., & Gardner, J. M. (2009). A pilot study of the Incredible Years Teacher Training program and a curriculum unit on social and emotional skills in community pre-schools in Jamaica. *Child: Care, Health and Development*, 35, 624–631. <http://dx.doi.org/10.1111/j.1365-2214.2009.00964.x>.
- Carlson, J. S., Tired, H. B., Bender, S. L., & Benson, L. (2011). The influence of group training in the Incredible Years Teacher Classroom Management program on preschool teachers' classroom management strategies. *Journal of Applied School Psychology*, 27, 134–154. <http://dx.doi.org/10.1080/15377903.2011.565277>.
- Conners, C. K., Sitarenios, G., Parker, J. D. A., & Epstein, J. N. (1998). Revision and restandardization of the Conners Teacher Rating Scale (CTRS-R): Factor structure, reliability, and criterion validity. *Journal of Abnormal Child Psychology*, 26, 279–291. <http://dx.doi.org/10.1023/A:1022606501530>.
- Gardner, F. (2008). Effective parenting interventions – Breaking the cycle of disadvantage by helping troubled families. In J. Gross (Ed.), *Getting in early: Primary schools and early intervention* (pp. 58–67). London, UK: The Smith Institute and Centre for Social Justice.
- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 38, 581–586. <http://dx.doi.org/10.1111/j.1469-7610.1997.tb01545.x>.
- Gottfredson, D. C., & Gottfredson, G. D. (2002). Quality of school-based prevention programs: Results from a national survey. *Journal of Research in Crime and Delinquency*, 39, 3–35. <http://dx.doi.org/10.1177/002242780203900101>.
- Herman, K. C., Borden, L. A., Reinke, W. M., & Webster-Stratton, C. (2011). The impact of the Incredible Years Parent, Child, and Teacher training programs on children's co-occurring internalizing symptoms. *School Psychology Quarterly*, 26, 189–201. <http://dx.doi.org/10.1037/a0025228>.
- Hutchings, J. (2012). Support for teachers around the world. *Better: Evidence-based Education*, 5(1), 18–19.
- Hutchings, J., Daley, D., Jones, K., Martin, P., Bywater, T., & Gwyn, R. (2007). Early results from developing and researching the Webster-Stratton Incredible Years Teacher Classroom Management training program in North West Wales. *Journal of Children's Services*, 2(3), 15–26.
- Hutchings, J., Gardner, F., & Lane, E. (2004). Making evidence-based interventions work in clinical settings: Common and specific therapy factors and implementation fidelity (Report No. 524). In C. Sutton, D. Utting, & D. Farrington (Eds.), *Support from the start: Working with young children and their families to reduce the risks of crime and antisocial behavior* (pp. 69–79). London, UK: Department for Education and Skills.
- Hutchings, J., Williams, M. E., Martin, P., & Pritchard, R. O. (2011). Introduction of the Incredible Years Child and Teacher programs in Gwynedd: Levels of behavioral difficulties among young welsh school children. *Welsh Journal of Education*, 15(1), 103–115.
- Jones, K., Daley, D., Hutchings, J., Bywater, T., & Eames, C. (2008). Efficacy of the Incredible Years program as an early intervention for children with conduct problems and ADHD: Long-term follow-up. *Child: Care, Health and Development*, 34, 749–756. <http://dx.doi.org/10.1111/j.1365-2214.2007.00747.x>.
- Kendall, P., & Wilcox, L. (1979). Self-control in children: Development of a rating scale. *Journal of Consulting and Clinical Psychology*, 47, 1020–1029. <http://dx.doi.org/10.1037/0022-006X.47.6.1020>.
- Leflot, G., Van Lier, P. A. C., Onghena, P., & Colpin, H. (2010). The role of teacher behavior in the development of disruptive behaviors: An intervention study with the Good Behavior Game. *Journal of Abnormal Child Psychology*, 38, 869–882. <http://dx.doi.org/10.1007/s10802-010-9411-4>.
- Martin, P. A., Daley, D., Hutchings, J., Jones, K., Eames, C., & Whitaker, C. J. (2010). The Teacher–Pupil Observation Tool (T-POT): Development and testing of a new classroom observation measure. *School Psychology International*, 31, 229–249. <http://dx.doi.org/10.1177/0143034310362040>.

- National Institute for Health Clinical Excellence (2006). Parent training/education programs in the management of children with conduct disorder (NICE technology appraisal guidance 102). Retrieved from http://www.guideline.gov/summary/summary.aspx?ss=15&doc_id=10313&nbr=5413
- National Statistics (2006). Schools in Wales: General statistics 2005 (first release, SDR 7/2006). Retrieved from <http://wales.gov.uk/docrepos/40382/40382313/statistics/schools-2006/sdr7-2006-e.pdf?lang=en>
- Office for National Statistics (2007). *Three years on: Survey of the development and emotional well-being of children and young people*. London, UK: Author Retrieved from <http://www.statistics.gov.uk/>.
- Raver, C. C., Jones, S., Li-Grinning, C., Metzger, M., Champion, K., Sardin-Adjei, L., et al. (2008). Improving preschool classroom processes: Preliminary findings from a randomized trial implemented in Head Start settings. *Early Childhood Research Quarterly*, 23, 10–26. <http://dx.doi.org/10.1016/j.ecresq.2007.09.001>.
- Reinke, W. M., Stormont, M., Webster-Stratton, C., Newcomer, L. L., & Herman, K. C. (2012). The Incredible Years Teacher Classroom Management program: Using coaching to support generalization to real-world classroom settings. *Psychology in the Schools*, 49, 416–428. <http://dx.doi.org/10.1002/pits.21608>.
- Rimm-Kaufman, A. E., Early, D. M., Cox, M. J., Saluja, G., Pianta, R. C., Bradley, R. H., et al. (2002). Early behavioral attributes and teachers' sensitivity as predictors of competent behavior in the kindergarten classroom. *Journal of Applied Developmental Psychology*, 23, 451–470. [http://dx.doi.org/10.1016/S0193-3973\(02\)0012-4](http://dx.doi.org/10.1016/S0193-3973(02)0012-4).
- Shernoff, E. S., & Kratochwill, T. R. (2007). Transporting an evidence-based classroom management program for preschools with disruptive behavior problems to a school: An analysis of implementation, outcomes, and contextual variables. *School Psychology Quarterly*, 22, 449–472. <http://dx.doi.org/10.1037/1045-3830.22.3.449>.
- Stone, L. L., Otten, R., Engels, R. C. M. E., Vermulst, A. A., & Janssens, J. M. A. M. (2010). Psychometric properties of the parent and teacher versions of the Strengths and Difficulties Questionnaire for 4- to 12-year-olds: A review. *Clinical Child and Family Psychology Review*, 13, 254–274. <http://dx.doi.org/10.1007/s10567-010-0071-2>.
- Webster-Stratton, C. (1999). *How to promote children's social and emotional competence*. London, UK: Paul Chapman Publishing.
- Webster-Stratton, C. (2003). *The Teacher Classroom Management Programme* (Rev. ed.) Washington, DC: Incredible Years Ltd.
- Webster-Stratton, C. (2005). The Incredible Years: A training series for the prevention and treatment of conduct problems in young children. In E. D. Hibbs, & P. S. Jensen (Eds.), *Psychosocial treatments for child and adolescent disorders: Empirically based strategies for clinical practice* (pp. 507–555) (2nd ed.). Washington, DC: American Psychological Association.
- Webster-Stratton, C., & Herbert, M. (1994). *Troubled families — Problem children: Working with parents: A collaborative process*. Chichester, UK: Wiley.
- Webster-Stratton, C., Mihalic, S., Fagan, A., Arnold, D., Taylor, T., & Tingley, C. (2001a). *The Incredible Years: Parent, Teacher and Child Training Series: Blueprints for violence prevention, book eleven*. Boulder, CO: Centre for the Study and Prevention of Violence, Institute of Behavioral Science, University of Colorado.
- Webster-Stratton, C., & Reid, M. J. (2002). The Incredible Years classroom management teacher training program: Content, methods, and process. Retrieved from <http://www.son.washington.edu/centers/parenting-clinic/opendocs/teachertrainingprogram.pdf>
- Webster-Stratton, C., Reid, M. J., & Hammond, M. (2001b). Preventing conduct problems, promoting social competence: A parent and teacher training partnership in Head Start. *Journal of Clinical Child and Adolescent Psychology*, 30, 283–302. http://dx.doi.org/10.1207/S15374424JCCP3003_2.
- Webster-Stratton, C., Reid, M. J., & Hammond, M. (2004). Treating children with early-onset conduct problems: Intervention outcomes for parent, child, and teacher training. *Journal of Clinical Child and Adolescent Psychology*, 33, 105–124. http://dx.doi.org/10.1207/S15374424JCCP3301_11.
- 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. <http://dx.doi.org/10.1111/j.1469-7610.2007.01861.x>.
- Webster-Stratton, C., & Taylor, T. (2001). Nipping early risk factors in the bud: Preventing substance abuse, delinquency, and violence in adolescence through interventions targeted at young children (ages 0–8 years). *Prevention Science*, 2, 165–192. <http://dx.doi.org/10.1023/A:1011510923900>.
- Weisz, J. R., Sandler, I. N., Durlak, J. A., & Anton, B. S. (2005). Promoting and protecting youth mental health through evidence-based prevention and treatment. *American Psychologist*, 60, 628–648. <http://dx.doi.org/10.1037/0003-066X.60.6.628>.
- Werner, E. E. (1999). How children become resilient: Observations and cautions. In N. Henderson, B. Benard, & N. Sharp-Light (Eds.), *Resiliency in action: Practical ideas for overcoming risks and building strengths in youth, families, and communities* (pp. 115–134). Gorham, ME: Resiliency in Action.
- Williford, A. P., & Shelton, T. L. (2008). Using mental health consultation to decrease disruptive behaviors in preschoolers: Adapting an empirically supported intervention. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 49, 191–200. <http://dx.doi.org/10.1111/j.1469-7610.2007.01839.x>.