Downloaded from UvA-DARE, the institutional repository of the University of Amsterdam (UvA)
http://hdl.handle.net/11245/2.73778

File ID	uvapub:73778
Filename	6: Supporting teachers' relationships with disruptive children : the potential of relationship-focused reflection
Version	unknown

SOURCE (OR PART OF THE FOLLOWING SOURCE):

Type	PhD thesis
Title	Relationships between teachers and disruptive children in kindergarten : an exploration of different methods and perspectives, and the possibility of change
Author(s)	J.L. Spilt
Faculty	FMG: Research Institute Child Development and Education (CDE)
Year	2010

FULL BIBLIOGRAPHIC DETAILS:
http://hdl.handle.net/11245/1.330667

Copyright

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content licence (like Creative Commons).

UvA-DARE is a service provided by the library of the University of Amsterdam (http://dare.uva.nl)
(pagedate: 2014-11-20)

6 Supporting teachers' relationships with disruptive children: The potential of relationship-focused reflection*

Guided by attachment research and Pianta's recommendations (1999), a relationship-focused reflection program (RFRP) was developed that targeted teachers' mental representations of relationships with specific children. The relative efficacy of the RFRP was examined in a randomized comparative trial with repeated measures. Thirty-two teachers were assigned to the RFRP or a comparison intervention aimed at changing teacher behavior. Per teacher, two children (N = 64) with above-median levels of externalizing behavior were selected. Multilevel growth modeling was used to explore intervention effects on teachers' perceptions of relationship quality (Closeness, Conflict, and Dependency), self-reported and observed pedagogical practices (Socioemotional Support, Behavior Regulation; Sensitivity, Behavior Management Quality). Teaching Efficacy was included as a moderator. The RFRP yielded changes over time in closeness for about half of the teacher-child dyads. In addition, teachers with high efficacy beliefs were more likely to report declines in conflict and increases in socioemotional support than low-efficacy teachers. Lastly, significant increases in observed sensitivity were found. These effects were different from those found in the comparison condition. Together, the study provided first support for the potential of reflection on specific relationships to promote teacher-child relationships.

* Spilt, J. L., Koomen, H. M. Y., Thijs, J. T., & van der Leij, A. Supporting teachers' relationships with disruptive children: The potential of relationship-focused reflection. *Submitted for publication*.

6.1 Introduction

Researchers increasingly consider relationships between teachers and students from an attachment perspective. It is believed that children can form attachment-like relationships with non-parental caregivers such as teachers, and also that these relationships foster children's school adjustment and learning (e.g., Hamre & Pianta, 2001; O'Connor & McCartney, 2007; Pianta, Steinberg, & Rollins, 1995; Pianta & Stuhlman, 2004). Especially in the early school years, teachers are recognized as secondary attachment figures who can provide care and support in stressful situations, and a secure base for exploration and learning (e.g., Koomen & Hoeksma, 2003). Therefore, researchers have called for the development of interventions that are specifically aimed at teacher-child relationships. In the current study, we evaluated an intervention designed to improve teachers' relationships with disruptive students. Attachment theory stresses the importance of relationship representations or working models. Hence, we examined the impact of targeting teachers' mental relationships models of specific children.

Problematic relationships are frequently observed between teachers and children who display non-compliant and disruptive behavior. Those relationships with disruptive children are often highly conflictual and elicit feelings of anger, distress, and helplessness in teachers (Spilt & Koomen, 2009). Teachers have been observed to be less sensitive and more controlling towards behaviorally-challenging children (Fry, 1983; Rimm-Kaufman et al., 2002; Stuhlman & Pianta, 2002). Poor relationships appear to exacerbate children's behavior problems, which may undermine children's learning potential and increase the risk of school failure (e.g., Doumen, Verschueren, Buyse, Germeijs, Luyckx, & Soenens, 2008; Ladd & Burgess, 2001). Conversely, when teachers succeed in forming close relationships with disruptive children, decreases in problem behavior are observed (e.g., Burchinal, Peisner Feinberg, Pianta, & Howes, 2002; Meehan, Hughes, & Cavell, 2003; Silver, Measelle, Armstrong, & Essex, 2005).

Unfortunately, the often hostile and disobedient behavior of children with externalizing problems makes it difficult for teachers to build positive relationships. Especially the heightened negative feelings of teachers are reason for concern. Teachers often appear to struggle with negative emotions when interacting with disruptive pupils (Hargreaves, 2000; Jennings & Greenberg, 2009;

Spilt & Koomen, 2009; Sutton, 2004). Negative feelings undermine teachers' commitment and responsibility taking, and limit coping skills and the ability to generate new strategies in dealing with children's behavior (Schutz, Hong, Cross, & Osbon, 2006; Sutton & Wheatley, 2003), eventually leading to exhaustion and burnout (e.g., Hargreaves, 1998; Montgomery & Rupp, 2005). Moreover, similar to what has been observed for mother-child interactions (Button, Pianta, & Marvin, 2001), negative feelings such as anger and frustration are conveyed in negative teacher behavior towards specific children (Stuhlman & Pianta, 2002). The quality of relationships between students and teachers appears also related to teachers' psychological functioning (Hamre, Pianta, Downer, & Mashburn, 2008; Mashburn, Hamre, Downer, & Pianta, 2006). For instance, teachers with low efficacy beliefs have been found to report higher levels of conflict with preschoolers than was expected based on the severity of disruptive behavior problems (Hamre et al., 2008).

Considering the difficulties for teachers to form secure relationships with disruptive children, the currently studied intervention addressed teachers' affective experiences with individual children that exhibit externalizing behavior. Such programs are critically needed as there seems to be little systematic attention to teachers' interpersonal experiences in teacher training and consultation (Hargreaves, 1998; Hughes, Cavell, & Jackson, 1999; Jennings & Greenberg, 2009; Ray, 2007; Schiff & BarGil, 2004; Williford & Shelton, 2008). As attachment theory has proven a valuable framework to study teacher-child relationships, it may also inform the development of interventions that target those relationships. One of the findings of attachment research is that the caregiver's capacity for reflection and understanding is vitally important for secure parent-child attachment (Slade, Grienenberger, Bernbach, Levy, & Locker, 2005). Attachment scholars have proposed the notion of reflective functioning (RF), which is the capacity to think about one's own and another's behavior in terms of underlying mental states such as feelings and intentions (Fonagy, Steele, Steele, Moran, & Higgitt, 1991). Whereas actual maternal behavior could not explain linkages between mother and infant attachment, there is some evidence that parental reflective functioning is the primary mechanism behind the intergenerational transmission of attachment (Slade et al., 2005; Van IJzendoorn, 1995). Consequently, it has been argued that intervention programs should help

mothers *think about and reflect on* their behavior rather than *change* behavior (Slade, 2006; Slade et al., 2005). First evaluations of such programs for substance-abusive mothers show promising results, suggesting that enhanced representations of caregiving and increased capacity for RF correspond with increases in sensitive behavior (Suchman, DeCoste, Castiglioni, Legow, & Mayes, 2008). In the same way, Pianta asserted that enhancement of teachers' mental models of relationships with specific children contributes to the dyadic relationship (cf. Pianta, 1999).

Analogous to mothers' representations of relationships with their children, teachers are believed to construct mental models of their relationships with students (Pianta, Hamre, & Stuhlman, 2003). Mental representations or internal working models are sets of internalized feelings and cognitions that guide expectations and behaviors toward others in a consistent and predictable manner and as such are central to interpersonal relationships (Bowlby, 1969; Bretherton, 1985). Teachers' representational models of a particular teacher-child relationship are thought to contain expectations and beliefs about the self as a teacher, about various roles (e.g., caregiver, disciplinarian, and instructor), and about the child in relation to the teacher (Pianta, 1999b; Pianta et al., 2003). Those models are constructed during numerous interactions with the child but are also believed to be, at least partly, rooted in teachers' own experiences as a teacher and of being taught and parented themselves. Representational models are believed to operate largely outside conscious awareness, and as a result teachers commonly have limited awareness of how their behavior is guided by underlying mental representations. Pianta (1999) describes how representations of self and others may reinforce themselves over time in a self-fulfilling way as people are inclined to seek information consistent with existing beliefs. In this way, mental representations can become highly stable and sometimes very constraining on teacher-child relationships. Therefore, rather than starting with changing teachers' behavior, Pianta (1999) contended that consultation should be directed first at constructing more flexible and differentiated representations of the relationship with a student through a reflective process.

To engage teachers in a reflective process, relationship narratives seem the obvious tool. In the field of teacher education narratives are presented as means to facilitate reflective practice and professional growth: *'...the value lies in the fact*

that it can serve as a basis for teachers' reflection, or, what is the same, it offers the possibility of going from the irrational to the rational, from unawareness to awareness, from the implicit to the explicit, from ignorance and custom to knowledge and reflection' (Clemente & Ramírez, 2008, p. 1257). Moreover, teachers themselves appear to enjoy such narrations and value them as a source of new insights (Pianta, 1999). Pianta proposed that when teachers narrate their representations, consultants could help summarize and label the narrations in more general terms guided by scientific theory, thereby increasing teachers' awareness and understanding. The Teacher Relationship Interview (TRI; Pianta, 1999) is developed to elicit such relationship narratives from teachers. The development of the TRI has been guided by a conceptual and methodological approach largely similar to other attachment-based interviews (Parent Development Interview; Button et al., 2001; Adult Attachment Interview; Main, Kaplan, & Cassidy, 1985). Importantly, research using attachment priming tasks suggested that especially narrative interview techniques are apt to tap into unconscious working models of attachment (Maier, Bernier, Perkrun, Zimmermann, & Grossmann, 2004). Because the TRI systematically attends to teachers' emotional states during salient interpersonal events, sources of emotional distress could be revealed and attended to. In the current intervention, we used the TRI to help teachers narrate and reflect on their relational experiences with specific children.

Present study

The aim of this research was to conduct a first evaluation of the effects of a relationship-focused reflection program (RFRP) for teachers. This program aimed to support teachers in building supportive relationships with disruptive children. It was designed to evoke changes in teachers' beliefs and orientations concerning their relationships and pedagogical approach towards individual students with above-median levels of externalizing behavior. Intervention effects were evaluated on the basis of learning as reflected by changes in teachers' perceptions and behavior. Outcome variables were: 1) teacher-reported relationship quality, 2) teacher-reported pedagogical approaches, and 3) observed quality of pedagogical practices of teachers. The first two measures were chosen because mental representations are believed to contain internalized beliefs and feelings about

relationships with specific children as well as professional scripts about appropriate pedagogical practices that are closely related to interpretations of students' behaviors and appraisals of relationships (Pianta et al., 2003; Spilt & Koomen, 2009; Thijs, Koomen, & Van der Leij, 2008). It was expected that teachers would report changes in relationship quality and pedagogical practices. Though at first sight declines in perceived relationship quality seem undesirable, such changes may not necessarily reflect a negative intervention effect. For example, a teacher could report more conflict in the relationship with a particular child not because of an actual increase in conflict but because the teacher may have become more aware of her own negative emotions and interactions with the child. This increase in conflict could then be viewed as a positive effect of the intervention because it reflects a shift from 'ignorance and custom' to 'acknowledgement and reflection' (Clemente & Ramírez, 2008). We included behavior observations because mental representations of relationships are believed to guide behavior of caregivers and the degree of responsiveness to a child's unique pedagogical needs. Accordingly, we predicted increases in the observed quality of pedagogical practices.

To test the hypothesis that intervention aimed at reflection is more powerful to enhance teacher-child relationships than attempts to change teachers' actual behavior, we examined the relative efficacy of the RFRP against a control intervention that was aimed at changing teachers' behavior. This was done in a cluster-level randomized comparative trial with repeated measures. The control intervention comprised a teacher training in caregiver-child interaction patterns labeled Interpersonal Skills Training (IST) that a) aimed at modification of teacher behavior, b) was based on scientific theory about interpersonal behavior, and c) could be delivered and tested using a similar design and delivery plan as the RFRP. Noteworthy, this intervention program did not focus on interactions of teachers with *specific* children. Based on attachment research, we expected the RFRP to evoke significantly stronger changes in teachers' perceptions and more pronounced increases in the quality of observed behavior than the IST.

Moderating effects of teaching efficacy beliefs were explored. We expected efficacy to be related to positive changes in relationship perceptions because previous studies have shown that high efficacy is associated with more positive

perceptions of teacher-child relationships (Hamre et al., 2008; Mashburn et al., 2006).

6.2 Method

Sample

Participants. The sample consisted of 32 teachers and 64 kindergarten children with externalizing behavior (45 boys) from 15 Dutch regular primary schools. Teachers were on average 40.0 years old (*SD* = 11.9), had 13.4 years experience in education (*SD* = 10.8), and worked 6.5 shifts (mornings and/or afternoons) per week in the same class (*SD* = 1.8). For each teacher, two children participated. Children's mean age was 66.9 months (*SD* = 5.3). Ethnical composition was 46 Dutch, 3 Turkish, 3 Moroccan, 7 Surinam, 2 other nationalities, and 3 unknown. Informed consent was obtained from parents and teachers.

Teachers and children participated in a larger research project and were selected from a sample of 188 kindergarten children (104 boys) and 47 teachers (all female) from 23 schools. Teachers were assigned to either the IST (16 teachers, 32 children) or RFRP condition (16 teachers, 32 children). Random assignment was conducted at the school level. No differences between the conditions were found with respect to teacher age, teaching experience, workdays per week, efficacy beliefs, children's age, children's gender, and mean levels of externalizing and internalizing behavior. One difference was found pertaining to ethnical composition. In the RFRP-intervention, 39.1% non-Dutch children were included against 7% in the IST-intervention.

Recruitment and selection procedure

Request letters were sent addressing both school principals and teachers. Participation of teachers was voluntary and was rewarded with a small fee of 20 Euros. Per teacher, four children were selected, with two of them included in the current study. The selection procedure was in line with the purposes of the larger research project that sought to examine children with heterogeneous behavior profiles. Teachers completed a behavior checklist (see Measures) for all 5-year-old children in their class prior to data collection. In every class, children were

categorized into four groups: 1) 'average children' with low scores on Externalizing and Internalizing Behavior; 2) 'inhibited children' with high scores on Internalizing but not on Externalizing Behavior; 3) 'disruptive children' with high scores on Externalizing but not on Internalizing Behavior; and 4) 'inhibited-disruptive children' scoring high on both Externalizing and Internalizing Behavior. Cut-off scores were 1.33 and 1.21 on a scale ranging from 1 to 4 for Internalizing and Externalizing Behavior, respectively. Those values were based on the median values derived from a large randomly-selected kindergarten sample (N = 1559). From each group, one child was randomly selected for participation.

In the current study, only the children from group 3 and 4 were included. Multiple paired t-tests were performed to evaluate the selection procedure. One-tailed tests were conducted with the significance level adjusted to .01 to avoid chance capitalization. Higher mean scores on Externalizing Behavior were found for the selected children from groups 3 and 4 in comparison to children from group 1 and group 2 ($3\text{-}1\Delta_{Mean}$=.72, t (47) = 9.197, p < .001; $4\text{-}1\Delta_{Mean}$= .55, t (47) = 8.521, p <.001; $3\text{-}2\Delta_{Mean}$= .68, t (47) = 8.758, p <.001; $4\text{-}2\Delta_{Mean}$= .51, t (47) = 7.522, p <.001). With respect to Internalizing Behavior, group 3 did not differ significantly from group 1 but had lower scores than group 2 ($3\text{-}1\Delta_{Mean}$= -.07, t (47) = 1.825, p = .04; $3\text{-}2\Delta_{Mean}$= -.54, t (47) = -9.307, p <.001). Conversely, group 4 differed significantly from group 1 but not from group 2 ($4\text{-}1\Delta_{Mean}$= .42, t (47) = 6.963, p <.001; $4\text{-}2\Delta_{Mean}$= -.05, t (47) = .835, p =.41). These results confirmed the selection purposes.

Teachers were neither informed about the selection procedure nor the experimental design of our study.

Intervention programs

Relationship-Focused Reflection Program (RFRP). A relationship-focused reflection program was developed to help teachers think about their relationship with an individual student. The design of the intervention was largely in line with Pianta's (1999) recommendations for intervention programs to be aimed at teachers' representations of relationships rather than actual behaviors. Important ingredients of the intervention were narration and reflection, with special attention for positive and negative emotions that teachers experienced in their daily work with a particular child.

The intervention comprised two blocks of two individual sessions with a consultant, with each block focusing on a different child. In the first session, teachers were helped to narrate their mental representation of the relationship with the target student using the Teacher Relationship Interview (TRI; Pianta, 1999a). In the second session, the teacher started with reflecting on one strength and one weakness in their actual interaction with the target child based on 3-5 minutes of video-recordings (see Measures: Observation of pedagogical practices). Consultants helped teachers to associate between their narrated representation and actual interactions with the child on the video. Consultants then presented teachers with a unique relational profile with strengths and weaknesses based on the teacher's narrative. Consultants and teachers discussed the profile together and could make adaptations to reach agreement on the profile. This profile was used as starting point for further reflection including the recognition of dissonances between actual practices and evidence-based alternatives, consideration of personal feelings for the child, identification of area(s) for improvement, and discussion of ideas for starting positive changes. In the third and fourth session, the same procedure was followed for the second child.

Consultants used the TRI coding manual to create relational profiles. The coding manual contains detailed descriptions of relational constructs: Sensitivity of Discipline (pro-active versus reactive, obedience-focused styles), Secure Base (socioemotional support versus insensitivity), Perspective Taking (understanding versus ignorance of child's individual needs), Intentionality (intentional practices versus absence of efforts to stimulate child's development), Helplessness (feelings of failure, sorrow, and giving up versus confidence in ability to establish change), Anger (anger and disapproval), Positive Affect (liking and satisfaction), Neutralizing Negative Affect (avoiding talking about negative feelings versus comfortable with diffuse emotions). The validity of the TRI has been evidenced by significant associations with teachers' questionnaire reports, observed teacher-child interactions, and by showing differences between typical and disruptive children (Spilt & Koomen, 2009; Stuhlman & Pianta, 2002).

The narratives and collaborative construction of a relational profile facilitated attunement to the unique strengths and weaknesses in each dyad, allowing for individualization of the consultation sessions. Some degree of standardization of the intervention was achieved by using a detailed manual for consultants.

Interpersonal Skills Training (IST). The Interpersonal Skills Training (IST; Thijs, 2005) is a training for professional educators based on the model of interpersonal communication of Leary (1957). Koomen and colleagues (2006) discussed the potential of this theoretical framework to understand and change teacher-child interactions. Teachers are trained to evaluate their interaction patterns with children along the orthogonal dimensions affiliation or proximity (opposition-cooperation) and directivity or influence (dominance-submission). They learn how to change unfavorable interactions patterns with students by changing their own behavior. This approach is especially useful for behavior modification of teachers.

Table 1 Time schedule of data collection and intervention delivery

		Questionnaires (QS)	Observations (OBS)	Intervention sessions (IS)
Phase 1	Week 1	QS_1	OBS_{1a}	
	Week 2			
	Week 3		OBS_1	
				IS_{1a}
	Week 4			IS_{1b}
Phase 2	Week 5	QS_2	OBS_2	
				IS_{2a}
	Week 6			IS_{2b}
Phase 3	Week 7		OBS_3	
	Week 8			
	Week 9	QS_3	OBS_4	

Note: The mean of OBS1a and OBS1b was used in the analyses.

Design overview and procedures

Delivery of interventions. The intervention sessions were delivered weekly after school time in a private setting in the school and took between 45-60 minutes. The RFRP was delivered by one of two consultants. Consultants held a master degree in developmental or clinical psychology. Three different consultants that both held master degrees in social sciences delivered the IST.

Data collection. The study started in early spring. Teacher-report questionnaires were administered together in a single booklet. Observations were conducted in the classroom. Data gathering and intervention sessions were planned according to a fixed schedule that covered 9 weeks and started after the selection procedure was completed (see Table 1). Questionnaires were administered at three occasions: preintervention (Phase 1), in-between first and second blocks of intervention (Phase 2), and postintervention (Phase 3). Video recordings were done biweekly at five occasions: twice prior intervention (Phase 1), once in-between first and second blocks of intervention (Phase 2), and twice postintervention (Phase 3). Intervention sessions were delivered in week 3 and 4 (block 1), and 5 and 6 (block 2). Due to practical reasons, absences, and school

Measures

Behavior checklist. Teachers completed the Behavior Questionnaire for Two- to Six-Year-Olds-Modified (BQTSYO-M; Goossens, Dekker, Bruinsma, & De Ruyter, 2000; Spilt, Koomen, Thijs, Stoel, & Van der Leij, 2009; Thijs, Koomen, De Jong, Van der Leij, & Van Leeuwen, 2004). The BQTSYO-M is a Dutch adaptation of the widely-used Preschool Behavior Questionnaire (PBQ; Behar, 1977). This screening measure comprises broadband scales for Externalizing (14 items; e.g., 'Bullies other children', 'A busy child') and Internalizing Behavior (15 items; e.g., 'Shy or timid towards other children', 'Easily upset'). Items are rated on a 4-point scale ranging from 1 (*absolutely not characteristic*) to 4 (*very characteristic*). Previous research on teacher-child relationships has supported the reliability and validity of the scale (Cronbach's alpha's $\geq$ .81 and .91 for Internalizing and Externalizing Behavior, respectively (Thijs & Koomen, 2009; Thijs et al., 2008).

Teacher-child relationships. The Student-Teacher Relationship Scale (STRS; Pianta, 2001) aims to measure teachers' perceptions of the relationship with a

particular student. A shortened version of the authorized Dutch translated and adapted version of the STRS was employed (Koomen, Verschueren, & Pianta, 2007). The questionnaire comprises three subscales: Closeness (6 items, e.g., 'I share an affectionate and warm relationship with this child'), Conflict (8 items, e.g., 'This child and I always seem to be struggling with each other') and Dependency (5 items, e.g., 'This child asks for my help when he/she really does not need help'). Closeness reflects the degree of warmth and open communication between the teacher and child. Conflict refers to negative and coercive teacher-child interactions, whereas Dependency represents levels of clinginess and overdependence of the child. Items are rated on a 5-point scale ranging from 1 (*not at all applicable*) to 5 (*highly applicable*). Adequate psychometric properties have been reported (Doumen, Verschueren, Koomen, & Buyse, 2008; Koomen et al., 2007). Internal consistency scores ranged between .88-93 for Closeness, .88-.91 for Conflict, and .75-.82 for Dependency in previous research, and stability coefficients at three measurement occasions during one school year ranged between .73-.84, .74-.83, and .67-.77, respectively (approximately 13 weeks). The validity has been supported by significant associations with classroom observations, teacher stress indices, and children's sociobehavioral adjustment and engagement. In the present study, Cronbach's alpha coefficients were .78, .86, and .75 for respectively Closeness, conflict, and dependency.

Self-reported pedagogical approach. The Teacher Pedagogical Practice Questionnaire (TPPQ; Thijs, Koomen, & Van der Leij, 2006) was employed to obtain teacher reports of pedagogical practices to individual children. The measure contains two subscales that are rated on a 5-point scale (1 = *absolutely not characteristic*; 5 = *very characteristic*). Socioemotional Support refers to teacher practices aimed at promoting the child's socioemotional well-being (e.g., 'More than other children I try to make this child feel safe' and 'I intervene when this child feels ill at ease'). Behavior Regulation encompasses teacher practices that indicate attempts to control and regulate the child's behavior through limit setting, behavioral reinforcement, and teaching social skills (e.g., 'I speak individually to this child about his/her social behavior' and 'I set clear limits to this child's behavior').

Previous research involving three independent samples of kindergarten has provided evidence for the psychometric strength of the scale (Thijs et al., 2006,

2008). Exploratory and confirmatory factor analyses supported the two-factor structure of the scale. In addition, unique associations have been found with teachers' free descriptions of their own pedagogical practices, teacher-child relationships, and teachers' appraisals of child behavior (i.e., Hindrance, Disruptiveness, Internalizing and Externalizing Behavior). Teachers reported more behavior regulative practices when they regarded children's behavior as under-controlled and disruptive. In contrast, they reported more socioemotional supportive actions when dealing with socially-inhibited children. These results suggest that the TPPQ measures a teacher's intended pedagogical approach or strategy in response to the perceived behavioral and emotional needs of a child. Both scales were found sufficiently reliable in the studies of Thijs and colleagues. For Socioemotional Support, consistency scores ranged between .76-.77, and for Behavior Regulation between .83-.85. In the current sample, Cronbach's alpha's of .73 and .84 were found, respectively.

Observed pedagogical practices. Video-recordings of teacher-child interactions were made during a dyadic-interaction task in a small-group setting in which the four selected children participated. This was done in the classroom during regular school days. There is evidence for the predictive validity of such structured observations above naturalistic observations (Zaslow et al., 2006).

The task consisted of series of pictures that reflected a sequence of events. Children had to place the pictures in a logical chronological order and tell and explain the story to the teacher. When teachers wanted to raise the level of difficulty, they could remove the last picture of a sequence and ask the child to come up with an alternative ending of the story. This interaction task has been successfully employed in previous research (Thijs & Koomen, 2008). The task took on average 15 minutes to complete. Other children not included in the study were present in the classroom during the small-group observation but worked independently. Teachers were informed that we were interested in mutual interactions and were encouraged to provide assistance as they were used to do. In case one of the four children was absent, another child with a similar behavior profile was chosen from the screening list for whom parents had provided informed consent. We did not include any data regarding this replacing child.

To measure teachers' sensitivity and behavior management quality, an adapted version of the Classroom Assessment Scoring System (CLASS) was employed

that has been successfully used in previous research (Buyse, Verschueren, Doumen, Van Damme, & Maes, 2008; Verschueren, Van de Water, Buyse, & Doumen, 2006). In contrast to the original CLASS (La Paro, Pianta, & Stuhlman, 2002), interactions between teachers and individual children were observed and not between teachers and the classroom as a whole or among children. The adapted Sensitivity scale measures the extent to which a teacher provides comfort, reassurance, and support with respect to a child's academic and socioemotional needs. Teachers receiving high scores were responsive to the child's needs and showed acceptance of the child's (negative) emotions such that they provided the child a secure base that enabled the child to take emotional and academic risks. The scale Behavior Management comprises teachers' ability to prevent and redirect a child's misbehavior. High scores indicated awareness of the child's behavior, brief and proactive responses, and clear communication of rules and expectations.

The scales were rated on a 7-point scale by independent coders who were unknown of the study purposes and procedures. The videotapes were rated in a random order. To examine interobserver-reliability, a random subsample was double coded (n = 72; 8.2%). Intraclass correlations (ICCs) for Sensitivity and Behavior Management in the total sample including all four children per teacher were .67 and .72, respectively. ICCs of .64 and .67 were found in the current subsample of children (i.e., group 3 and 4). Those ICCs indicate adequate interobserver agreement, and suggest that only a small proportion of total variance is attributable to rater effects (Cicchetti et al., 2006).

Teaching efficacy. The Teacher Efficacy Scale (TES; Gibson & Dembo, 1984) is a 30-item teacher-report measure of teacher efficacy. Adapted forms of this scale have been widely used, and in particular the validity of the personal teaching efficacy items has been supported (for a review, see Henson, 2002). Personal teaching efficacy reflects the degree to which teachers believe that their own capabilities influence children's learning and behavior (e.g., 'When a student does better than usual, many times it is because I exerted a little extra effort.'). In the current study, a 17-item version was administered (Deemer & Minke, 1999). Because the questionnaire was presented in a single booklet together with the STRS, a 5-point scale was used to ease completion (1 = *totally disagree*, 5 = *totally agree*). In line with the research of Deemer and Minke (1999), explorative

factor analysis on the current sample yielded a one-factor solution representing primarily personal teaching efficacy. The internal consistency of the scale was adequate (α =.73).

Data analyses

Conventional growth modeling. The repeated measures allowed for the assessment of intervention effects on trajectories of change instead of focusing on intervention effects at a specific point in time. In a multilevel framework, we analyzed linear growth curves using MLwiN version 2.02 (Rasbash, Browne, Healy, Cameron, & Charlton, 2005). Three-level random intercepts models were estimated with measurement occasions (level 1) nested within children (level 2), and children nested within teachers (level 3). As intervention effects could depend on teacher and child characteristics, we expected variations between dyads in change rate and/or direction of change (i.e., random slope variance components). Therefore, separate slopes were estimated and compared for the IST and RFRP groups in order to examine the significance of random variance in each condition. We sought to explain random slope variance at level 3 by testing cross-level interactions between slopes and the time-invariant covariate Efficacy (grand-mean centered). Other covariates that were explored were teacher age, teacher experience, intervention satisfaction, and externalizing and internalizing child behavior. Because no significant results were found, those analyses were not reported.

Data were analyzed under the assumption missing at random (MAR). ML estimation was employed using iterative generalized least squares (IGLS). To test the significance of the estimated parameters (i.e., fixed effects), the Wald-test was performed (two-sided). For the assessment of between-group differences in slopes and random slope effects, the chi-square deviance test was used (two-sided). Considering the sample size, the level of significance was set at .10 in order to reduce the chance of Type-II error. Residual plots were inspected to check for normal distribution of the residuals and homoscedasticity (i.e., residual variance is independent of the values of predictors). The impact of extreme scores on the models was evaluated based on influence values that were multilevel equivalents of the DFITS measure based on residuals and leverage values (Bollen & Jackmam, 1985; Langford & Lewis, 1998).

Latent Class Growth Analysis. Conventional growth modeling estimates a growth curve under the assumption that one mean growth curve adequately describes growth for all individuals from a single population. Alternately, latent class growth analysis (LCGA; Nagin, 1999) assumes a mixture of distributions representing unobserved groups (i.e., latent classes) of individuals that follow different slope trajectories. This approach is helpful to evaluate intervention success because, rather than assuming equal intervention effects on all participants, there may be unobserved subgroups of individuals that are differentially impacted by the intervention (i.e., one group of individuals improves, while another remains unchanged). LCGA can capture this heterogeneity in intervention response (Lennon, McAllister, Kuang, & Herman, 2005; Muthén et al., 2002).

LCGA was conducted in the Mplus program version 4.2 (Muthén & Muthén, 1998-2006) following guidelines provided by Jung and Wickrama (2008). The hierarchical structure of the data was taken into account using the cluster sampling option. To determine the numbers of classes, models estimating different k classes are run that are not nested within $k + 1$ class models. The Bayesian information criteria (BIC) value was used for model selection. These analyses were conducted separately in the RFRP and IST condition as recommend by Muthén and colleagues (2002).

6.3 Results

Preliminary analyses

Descriptives. Table 2 presents descriptive statistics. The dependent variables showed normal distributions as indicated by coefficients of skewness (Skw) and kurtosis (Krt) below .982 and 1.447, respectively. Correlations between variables were in the expected direction (see Table 3).

Missing values. Table 2 reports the number of missing values. Mostly, data were missing because teachers did not return the booklet at a certain measurement occasion, which resulted in missing data on all five questionnaire variables. Missing scores in the observations were due to absence during the small-group observation generally caused by illness of the child. No mean differences between

conditions were observed with respect to missing values in the questionnaire data (t (62) = .389, p >.10) and observations (t (62) = 1.414, p >.10). Children with missing values were compared to children with complete data by means of t or chi-square tests. No systematic differences were found on gender, ethnicity, age, or on the study variables (p >.10).

Extreme cases. Data were screened on multivariate outliers based on inspection of scatter plots of relationships between the dependent variables and casewise-diagnostics. Extreme values (Std. Residual > 2.5) were removed (see Table 2).

Pre-intervention differences. Independent t-tests indicated no pre-intervention differences between the RFRP and IST on Closeness (t (118) = -.203, p >.10), Conflict (t (118) = -.167, p >.10), Dependency (t (118) = -.620, p >.10), Socioemotional Support (t (118) = -.219, p >.10), Behavior Regulation (t (118) = .064, p >.10), observed Sensitivity (t (62) = .483, p >.10), observed Behavior Management Quality (t (62) = -1.464, p >.10), Externalizing Behavior (t (127) = .698, p >.10), Internalizing Behavior (t (127) = .577, p >.10), Teacher Age (t (30) = .592, p >.10), Teaching Experience (t (30) = .724, p >.10), and Teaching Efficacy (t (29) = -.401, p >.10).

Table 2 Descriptives of study variables (N = 64 children; 16 teachers)

Variables	Time 1 *M (SD)*	Time 2 *M (SD)*	Time 3 *M (SD)*	Time 4 *M (SD)*	Missing *n* (%)	Extremes *n* (%)
Teacher reports						
Closeness	4.09 (.70)	4.14 (.65)	4.15 (.67)		7 (3.6%)	0 (0.0%)
Conflict	2.31 (.95)	2.23 (.91)	2.24 (.91)		7 (3.6%)	2 (1.0%)
Dependency	2.30 (.74)	2.38 (.72)	2.25 (.73)		7 (3.6%)	3 (1.6%)
Socioemotional Support	2.93 (.78)	3.00 (.75)	2.85 (.86)		7 (3.6%)	1 (0.5%)
Behavior Regulation	3.51 (.89)	3.42 (.92)	3.40 (.89)		7 (3.6%)	0 (0.0%)
Externalizing	1.78 (.55)				0 (0.0%)	0 (0.0%)
Internalizing	1.39 (.41)				0 (0.0%)	0 (0.0%)
Efficacy	3.84 (.43)				1 (3.1%)	0 (0.0%)
Observations						
Sensitivity	5.10 (.77)	5.07 (.87)	5.16 (.85)	5.15 (.86)	16 (6.3%)	6 (2.3%)
Behavior Management Quality	5.17 (.87)	5.21 (.89)	5.25 (1.07)	5.18 (.88)	16 (6.3%)	3 (1.2%)

Table 3 Correlations between study variables at Time 1

Variables	1	2	3	4	5	6	7	8	9	10
Teacher reports										
1 Closeness	-									
2. Conflict	-.48**	-								
3. Dependency	.00	.02	-							
4. Socioemotional Support	-.05	-.07	.33*	-						
5. Behavior Regulation	-.47**	.56**	.06	.23[a]	-					
6. Externalizing	-.42**	.58**	-.18	-.02	.45**	-				
7. Internalizing	-.09	-.13	.23[a]	.41**	-.09	-.01	-			
8. Efficacy	.19	-.36**	-.09	.34**	-.07	.01	.02	-		
Observations										
9. Sensitivity	.10	-.12	.10	.10	.01	-.03	-.01	.12	-	
10. Behavior Management Quality	.33**	-.39**	-.06	.14	-.23[a]	-.24*	.17	.16	.57**	-

Note: *$p < .05$, **$p < .01$, [a]$p < .10$ (two-tailed)

Growth modeling

Conventional Growth Modeling (CGM) proceeded in four steps following guidelines of Hox (2002). First, unconditional models with two linear slopes (i.e., fixed effects) for the IST and RFRP group, respectively, were examined to indicate change over time in the independent variable in each condition. Second, to test whether the slopes of IST and RFRP were significantly different, the slopes were constrained to be equal and the decrease in model fit was evaluated using the chi-square deviance test. Third, random slope components were examined on a variable-by-variable basis to detect systematic differences in intervention effects across teacher-child dyads. When significant slope variation was found across teachers, Efficacy was added in the fourth step to explain different rates of change. Table 4 and 5 present the final multilevel models. Latent Class Growth Analyses were reported only in case of unexplained random slope variance.

Closeness. For Closeness, no group-mean change over time was found in the RFRP (β = .034, Z = 0.791, p > .10). The estimated slope in the RFRP did not differ significantly from the slope found in the IST, which was also not significantly different from zero (β = .030, Z =0.698, p > .10). When slopes were allowed to vary across individuals, a random variance component was found in the RFRP but not in the IST. The RFRP slope varied across children (Δx^2 = 10.918, Δdf = 2, p < .01) and teachers (Δx^2 = 6.655, Δdf =2, p < .05), suggesting systematic differences between teacher-child dyads. Efficacy, however, could not explain this random slope variance (p > .10).

LCGA was employed to gain more insight in the systematic variation in intervention effects in the RFRP. A model comprising three classes showed the best fit (two-class BIC = 176; three-class BIC = 171; four-class BIC = 181). Classification quality was good (Entropy = .89; Posterior Probability = .89-.99). Fifteen dyads showed a high and stable pattern of Closeness (intercept = 4.48, p < .001; slope = .06, p > .10), six dyads with low Closeness prior to intervention showed a declining pattern (intercept = 3.37, p < .001; slope = -.24, p < .05), and eleven dyads with low Closeness prior to intervention showed an increase following intervention (intercept = 3.63, p < .001; slope = .21, p < .10). Thus change was established for all children who started with poor levels of Closeness prior to intervention. The trajectories of Closeness are depicted in Figure 1.

Table 4 Multilevel regression models for relationship perceptions

	Closeness	Conflict 1	Conflict 2	Dependency
	B (SE)	*B (SE)*	*B (SE)*	*B (SE)*
Fixed part				
Intercept (β_0)	4.08 (.09)**	2.29 (.12)**	2.29 (.11)**	2.32 (.09)**
Slope IST (β_1)	.03 (.04)	-.09 (.05)[a]	-.07 (.04)	-.09 (.06)
Slope RFRP (β_2)	.03 (.06)	.03 (.07)	.06 (.05)	.02 (.06)
Efficacy (β_3)			-.55 (.13)**	
Efficacy x Slope RFRP (β_4)			-.17 (.10) [a]	
Random part				
Level 1 (occasions)				
$\sigma_e^2{}_0$	.090	.125	.117	.211
Level 2 (children)				
$\sigma_u^2{}_0$	.266	.728	.697	.310
$\sigma_u^2{}_1$	.000	.000	.000	.000
$\sigma_u^2{}_2$	.045	.000	.000	.000
$\sigma_u^2{}_4$	-	-	.217	-
Level 3 (teachers)				
$\sigma_v^2{}_0$	.068	.000	.000	.000
$\sigma_v^2{}_1$	.000	.000	.000	.000
$\sigma_v^2{}_2$	.023	.054	.000	.000
Deviance	*258.6*	*338.2*	*293.8*	*337.5*

Note 1: *$p < .05$,**$p < .01$, [a]$p < .10$

Note 2: Unstandardized β coefficients (*B*) are reported; Covariances between residual error terms (random part) were set free but not reported

Table 5 Multilevel regression models of teacher-reported and observed pedagogical practices

	Se Support 1	Se Support 2	Behavior R	Sensitivity	Behavior MQ
	B (SE)	*B (SE)*	*B (SE)*	*B (SE)*	*B (SE)*
Fixed part					
Intercept (β_0)	3.02 (.12)**	3.01(.12)**	3.51 (.11)**	5.08 (.10)**	5.21 (.10)**
Slope IST (β_1)	-.08 (.05)	-.09 (.06)	-.07 (.06)	-.03 (.05)	.03 (.06)
Slope RFRP (β_2)	-.07 (.08)	-.08 (.08)	-.05 (.05)	.09 (.055)[a]	.00 (.06)
Efficacy (β_3)		.14 (.18)			
Efficacy x Slope RFRP (β_4)		.39 (.17)*			
Random part					
Level 1 (occasions)					
$\sigma_e^2{}_0$	.140	.147	.194	.559	.706
Level 2 (children)					
$\sigma_u^2{}_0$	.304	.318	.621	.032	.030
$\sigma_u^2{}_1$	.000	.000	.000	.000	.000
$\sigma_u^2{}_2$	.036	.023	.000	.000	.000
Level 3 (teachers)					
$\sigma_v^2{}_0$	.237	.192	.000	.084	.035
$\sigma_v^2{}_1$	.000	.000	.000	.000	.000
$\sigma_v^2{}_2$	.057	.066	.000	.000	.000
Deviance	*334.7*	*321.5*	*367.0*	*560.6*	*608.0*

Note 1: $*p < .05$, $**p < .01$, $^a p < .10$

Note 2: Se = Socioemotional, R=Regulation, MQ=Management Quality

Note 3: Unstandardized β coefficients (B) are reported; Covariances between residual variance terms (random part) were set free but not reported

Conflict. No group-mean change in Conflict was found in the RFRP (β = .024, Z = 0.489, p > .10). In contrast, a decrease in Conflict was shown for the IST (β = -.092, Z = 1.80, p < .10). When the slopes in the IST and RFRP were constrained to be equal, the difference between the slopes appeared significant (Δx^2 = 2.825, Δdf = 1, p < .10). When slopes were allowed to vary across individuals, a random variance component at the teacher-level was found in the RFRP (Δx^2 = 8.941, Δdf = 1, p < .01). Efficacy was added to the model in order to explain this random slope (Δx^2 = 44.374, Δdf = 4, p < .001). A negative main effect was found on Conflict (β = -.549, Z = -4.19, p < .001) as well as cross-level interaction with RFRP-slope (β = -.169, Z = -1.78, p < .10), which explained all the variance in change rate between teachers. This interaction indicated that higher levels of teaching efficacy predicted decreases in conflict. Figure 2 depicts the interaction effect between slope and Efficacy on Conflict.

Dependency. For Dependency, no significant slopes were found in both IST (β = .085, Z = 1.518, p > .10) and RFRP (β = .020, Z = 0.357, p > .10), nor were random slope variance components detected (p > .10).

Socioemotional Support. No average change was found in Socioemotional Support for the IST (β = .076, Z = 1.407, p > .10) or RFRP (β = .075, Z = 1.377, p > .10). However, for the RFRP, the slope appeared to vary across children (Δx^2 = 7.861, Δdf = 2, p < .05) and teachers (Δx^2 = 7.575, Δdf = 2, p < .05). Efficacy contributed significantly to the model and partly explained random variance (Δx^2 = 13.176, Δdf = 2, p < .01). No main effect was found (β = .215, Z = 0.77, p > .10) but the interaction between Efficacy and RFRP-slope was significant (β = .385, Z = 2.305, p < .05), such that higher levels of Efficacy were associated with increases in Socioemotional Support (see Figure 3).

LCGA was performed to increase understanding of the random slope variance in the RFRP. A two-class model fitted the data best (one-class BIC = 253; two-class BIC = 229; three-class BIC = 231), and showed adequate classification quality (Entropy = .86; Posterior Probability = .95-.97). Teachers reported modest and stable levels of Socioemotional Support for about two-third of the children (n = 20; intercept = 2.40, p < .001; slope = .07, p > .10). For the other children, teachers reported high declining levels of Socioemotional Support (n = 12; intercept = 3.86, p < .001; slope = -.18, p < .05).

Behavior Regulation. For Behavior Regulation, no significant change was

found in both IST (β = -.070, Z = 1.272, p > .10) and RFRP (β = -.046, Z=0.852, p >.10). In addition, no significant random slope components were detected (p > .10).

Observed Sensitivity. Observations were conducted at five occasions of which two prior to intervention. The mean of the first two occasions was used as a pretest at T=0. Thus slopes were estimated across four time points. An increase in Sensitivity was shown for the RFRP (β = .093, Z = 1.69, p < .10) but not for IST (β = -.033, Z=0.622, p >.10). When the slopes were constrained to be equal, model fit decreased significantly, which indicated that the rate of change was different in the two conditions (Δx^2 = 3.654, Δdf = 1, p = .056). Random slope variance was zero in both conditions; therefore no cross-levels interactions were tested nor did we perform LCGA.

Observed Behavior Management. No significant change was found in Behavior Management Quality in either IST (β = .034, Z = 0.596, p > .10) or RFRP (β =-.002, Z = 0.003 p > .10) nor was random slope variance detected (p > .10).

Figure 1 Latent class trajectories of Closeness for teacher-child dyads in the Relationship-Focused Reflection Program (RFRP)

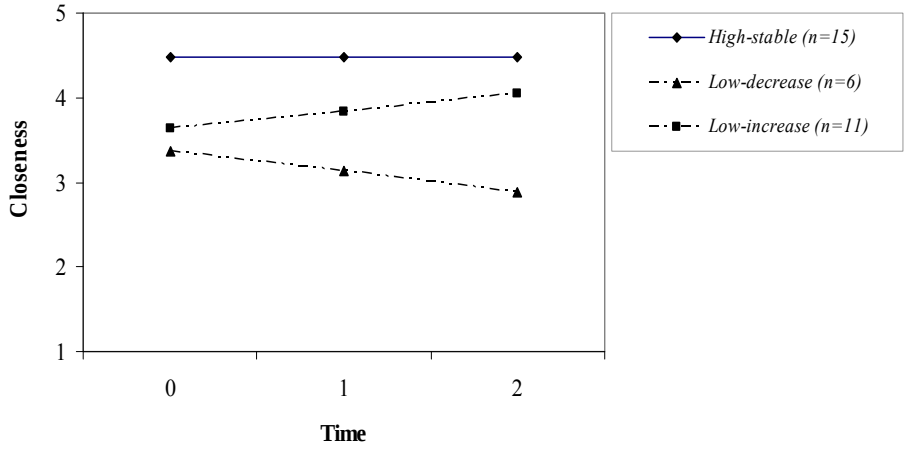


Figure 2 Efficacy as a moderator of changes in Conflict for teacher-child dyads in the Relationship-Focused Reflection Program (RFRP)

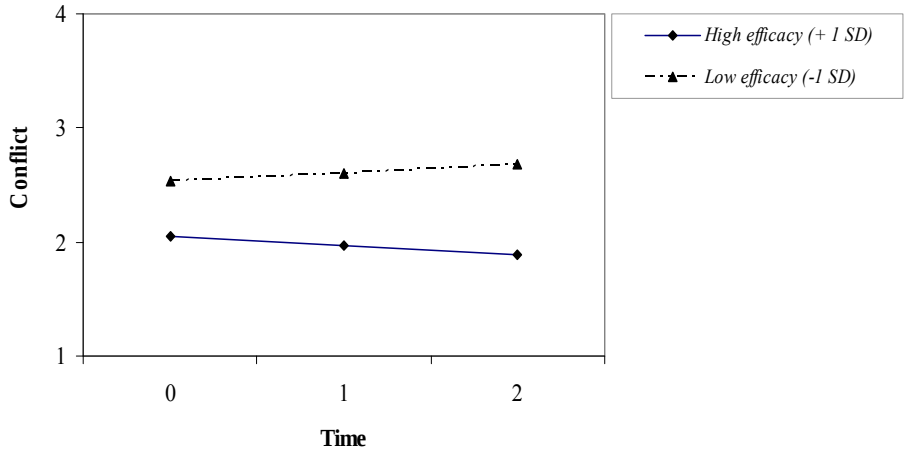
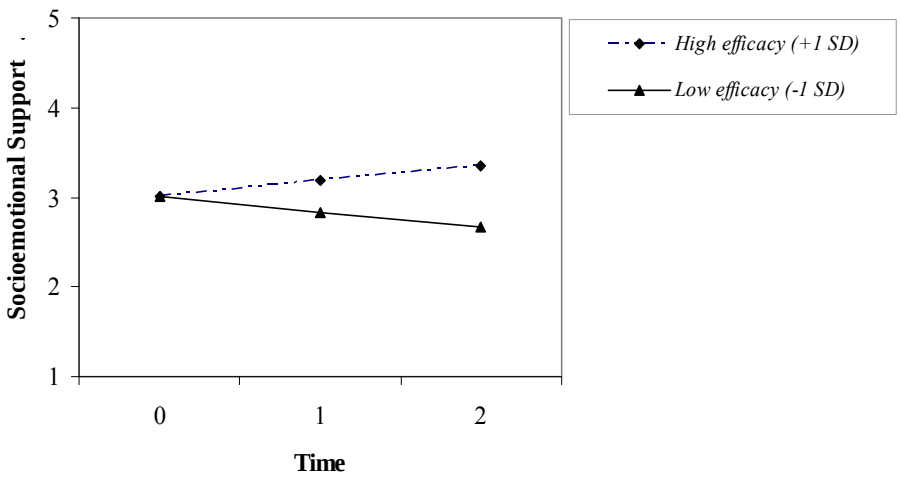


Figure 3 Efficacy as a moderator of changes in Socioemotional Support for teacher-child dyads in the Relationship-Focused Reflection Program



6.4 Discussion

This study is among the first to address the repeated call for interventions specifically aimed at teacher-child relationships. Guided by attachment theory and Pianta's ideas about teacher consultation (1999), we developed a relationship-focused reflection program (RFRP) to promote teachers' relationships with behaviorally-challenging children. The potential of the RFRP was supported by changes in perceived closeness for about half of the teacher-child dyads, and an increase in observed sensitivity across the whole group. In addition, differential intervention effects on conflict and intended provision of socioemotional support could be explained by teaching efficacy beliefs. Overall, the results were considerably different from those found for teachers receiving training aimed at behavior modification, and supported the relative efficacy of in-depth reflection on specific relationships.

In the RFRP condition, no group-mean changes were established in teacher-reported outcomes. However, systematic differences across teacher-child dyads were detected in intervention effects over time on closeness, conflict, and intended provision of socioemotional support. This suggests that subgroups of teachers did report changes on those outcome variables. For closeness, the modeling of latent trajectory classes nicely advanced understanding of these random effects. For about half of the dyads in the RFRP condition, a stable trajectory was detected with high closeness prior and post intervention, which seemed to suggest a ceiling effect. The majority of dyads with relatively non-close relationships at the start of the intervention showed increases in closeness. For a small group of children with relatively non-close relationships, however, teachers reported decreases in closeness. For those teachers, reflection on the interpersonal relationship and affective experiences may have strengthened a sense of relational difficulties. Though a decline in closeness is a reason for concern, it could represent a shift from ignorance to increased awareness of a lack of trust and warmth in the relationship with the child. Positive effects of such increased awareness could perhaps be that a teacher becomes more sensitive to the child's need for relatedness or that it may persuade a teacher to seek help from school mental health services. Future research is needed to answer the question why positive changes were observed for some and negative changes for others.

The detection of changes in closeness was in particular valuable because closeness is believed to reflect the teacher's and child's ability to social-emotionally connect with each other (Silver et al., 2005). Following the attachment framework, closeness especially can be interpreted as the extent to which the teacher functions as a source of security for a specific child. It's therefore interesting that changes in closeness were observed in all dyads with relatively low levels of closeness. Close relationships have the potential to buffer children against academic risks associated with social-behavioral problems, and insight in how to affect closeness is therefore highly valuable (Burchinal et al., 2002; Ladd & Burgess, 2001; Meehan et al., 2003; Silver et al., 2005).

Changes in conflict and provision of socioemotional support following relationship-focused reflection appeared dependent on efficacy beliefs of teachers. Prior to intervention, teachers with low efficacy beliefs tended to report more conflict and were less inclined to provide socioemotional support in relationships with disruptive children than teachers with high efficacy beliefs (see Table 3; Hamre et al., 2008; Mashburn et al., 2006). As anticipated, the RFRP program seemed to enlarge these differences as low efficacy was related to moderate increases in conflict and declines in socioemotional support. For low-efficacy teachers, reflection and increased awareness may be stressful or discouraging because teachers could be pessimistic about their ability to influence the relationship. Conversely, for teachers who hold strong beliefs about their ability to influence children's behavior and development, increased awareness and understanding of the relationship seemed to enlighten their views on relational conflict and strengthen beliefs about the importance of being socioemotionally supportive. Those changes in socioemotional support suggested a reappraisal of the socioemotional needs of the child following in-depth reflection (cf., Thijs et al., 2008). Importantly, improvements in sensitive behavior were observed for all teachers following relationship-focused reflection. This indicates that both low- and high-efficacy teachers profited in some way from the intervention.

Efficacy beliefs could not explain all systematic variance in changes in socioemotional support. To gain more insight in differential change across subgroups of teacher-child dyads, we modeled latent trajectories classes (LCGA). Somewhat surprisingly, those analyses did not indicate a subgroup of teachers showing an increasing pattern in socioemotional supportive practices, which

could have been expected for high-efficacy teachers, based on the conventional growth models (CGM). Two explanations could be considered. First, LCGA uses both intercept and slope differences to identify trajectory classes and, hence, estimates more parameters. Therefore, there may have been insufficient power to identify increasing trajectories. Second, differences could have risen because, unlike with the CGM analyses, we conducted LCGA for the RFRP group separately as recommended by Muthén and colleagues (2002). Unfortunately, we could not test the effects of Efficacy on latent trajectory classes because of the increased complexity of multilevel LCGA in relation to the limited sample size.

The RFRP induced group-mean changes (i.e., increases) in observed teacher sensitivity. This small but significant improvement in observed sensitivity was promising, especially in light of the generally observed declines in teacher sensitivity over time in samples with behaviorally-challenging children (Fry, 1983). Moreover, emotionally supportive teacher practices can protect at-risk children against academic underachievement and relational conflict (Hamre & Pianta, 2005).

RFRP did neither impact teacher-reported dependency and behavior regulation, nor the observed quality of behavior management. The first finding may not be surprising as Dependency appeared scarcely related to the TRI in previous research (Spilt & Koomen, 2009). Behavior management and compliance issues, however, are often salient themes in teachers' experiences with disruptive children (Stuhlman & Pianta, 2002). Probably because of this focus on behavior regulation, teachers tend to be less responsive to prosocial behavior, provide less sustained feedback, and show less attempts to engage children in learning activities (Fry, 1983; McComas, Johnson, & Symons, 2005). In this light, it is interesting that relationship-focused reflection did neither seem to affect teachers' intended provision of behavior regulation nor their actual behavior management quality but, instead, teachers tended both to change their provision of socioemotional support and to respond gradually more sensitive to the socioemotional and academic needs of disruptive children. Perhaps, reflection helped teachers shift their attention from compliance issues to children's emotional and academic lives.

Noteworthy, whereas teacher reports of relationship quality and pedagogical approaches were clearly dyadic measures, our observations of the actual quality

of pedagogical practices appeared not. Though pedagogical practices of teachers were observed for each child separately, the observer ratings contained limited variance at the child level. One reason could be the homogenous nature of the sample, which was a subset of disruptive children who were all targeted in the intervention. In the total sample from which those children were drawn, the proportions of variance at the child-level were higher and largely consistent with previous research (Buyse, 2007), which confirms the validity of our observations. It could be inferred that the observations of pedagogical practices showed primarily a teacher characteristic; that is a pedagogical style towards disruptive children.

The effects of the RFRP were evaluated in comparison to an intervention labeled Interpersonal Skills Training (IST). The IST comprised a more general approach (i.e., not focused on specific children) targeting teacher behavior in dyadic interactions. Teachers who participated in the IST reported significant declines in conflict. No effects were found on other teacher-reported outcomes nor were improvements in pedagogical practices observed. Furthermore, no systematic differences in intervention effects across teacher-child dyads were revealed. Together, the results suggested stronger effects of the RFRP on both teacher-reported and observed outcomes in comparison to the IST. In line with attachment research, we argue that the RFRP showed more potential because it helps teachers think about and reflect on dyadic relationship experiences rather than aiming at behavior modification. Yet, another explanation could be that the IST was less successful because it did not explicitly focus on teachers' interactions with specific children.

Noteworthy, the results were obtained in a sample of kindergarten children with mostly mild behavior problems. In contrast to randomly selected samples, externalizing scores were normally distributed. Mean levels of Closeness appeared largely similar to those of a large Dutch normative sample, whereas levels of Conflict were higher (Koomen et al., 2007). This pattern is in line with previous research (Spilt & Koomen, 2009). The distribution of scores on Closeness suggested a ceiling effect leaving little room for improvement for about half of the sample. Future studies could use samples with more serious conduct problems, for instance, including children whom teachers identify as 'problem students'. In addition, to examine long-term intervention effects and effects on

children's school adjustment, the inclusion of outcome measures of child functioning and repeated follow-up assessments should be considered in future evaluations of the RFRP.

Several limitations should be considered. First, the sample size was modest and especially the modeling of latent trajectory classes (LCGA) should be considered preliminary. Additionally, by increasing the alpha level to .10, we aimed to reduce the probability of a Type-II-error. In doing so, the likelihood of a Type-I-error was increased, which is accepting the RFRP as an effective intervention when it is not (i.e., false positive conclusion). Second, there was not a no-treatment control condition. Though it is possible that teachers' relationships with disruptive children improved through time alone, longitudinal studies have indicated significant increases in relational conflict and declines in teacher sensitivity across the course of a school year (Doumen et al., 2008; Fry, 1983). The advantage of the current study was that the comparison condition comprised an intervention that was similar in its delivery plan and design, and differed in content only. In doing so, it could be demonstrated that intervention effects did go beyond the simple effects of receiving attention or the expectation of change. Third, teachers participated voluntarily. Fourth, observer scores should be interpreted in the context of a specific small-group task activity. Though this is considered an adequate natural context to study dyadic interactions, it provides an incomplete assessment of teacher practices as interpersonal behavior could differ across contexts (i.e., situation variance, see Lakes & Hoyt, 2008). Moreover, the observations contained much variance at the lowest level (i.e., occasion variance). This suggests that the quality of pedagogical practices of teachers was driven by time-specific contextual influences that occur at the moment of observation (Lakes & Hoyt, 2008). High occasion variance may also suggest measurement error. This could be a concern because instability over time limits the detection of changes over time induced by the interventions. Final issues that may be considered are the length and timing of intervention. In the literature, there is no consensus regarding the optimal length of intervention. Whereas Pianta argues for a long-term focus to improve relationships (1999), attachment intervention research supports the efficacy of short-term preventive programs (Van IJzendoorn, Juffer, & Duyvesteyn, 1995). In addition, intervention delivered in the fall could have yielded stronger results as relationships may be more changeable earlier in

the school year.

In sum, this study established the relative efficacy of a dyad-focused intervention comprising relationship-oriented reflection for supporting relationships between teachers and disruptive children. The intervention yielded changes in both teacher perceptions and observed practices. This suggests that teachers reorganized their mental representation and obtained new insights after in-depth reflection that affected their actual responsiveness to children's needs. According to attachment research, these processes could be keys to improve security in adult-child relationships (e.g., Slade et al., 2005).

This study provided initial support of Pianta's idea to facilitate teachers' reflection on their interpersonal and emotional experiences with individual children through helping them narrate their relationship experiences and to label and 'mirror' those experiences (1999). The approach closely fits with the contemporary emphasis on the consulting role of school psychologists and can be easily integrated in existing teacher consultation models. School mental health professionals are encouraged to use the TRI as a relational framework to explore, understand, and adapt to teachers' relational difficulties with problem students. The TRI could provide starting points to set up consultation and engage teachers in a reflective process (Koomen, Verschueren, & Thijs, 2006; Stuhlman & Pianta, 2002).

In conclusion, the results supported the potential of in-depth reflection to promote relationships between teachers and children. The intervention affected teachers' relationship appraisals and pedagogical approach towards children with externalizing behavior as well as their observed responsiveness to the socioemotional and academic needs of those children.