CrossMark

Effects of injustice in primary schools on students' behavior and joy of learning

Natalie Ehrhardt-Madapathi[1] · Johanna Pretsch[2] · Manfred Schmitt[2]

Abstract What are the long-term effects of justice and injustice in school? This longitudinal study examined the bidirectional relationships of classroom justice with students' behavioral problems, well-being, and joy of learning. Using systematic observation on the singlechild level, teachers and external observers made high and low inference justice ratings of 245 primary school students. The justice ratings measured two aspects of classroom justice: pedagogical practices and teacher–child interactions. Results indicate that there is a bidirectional relationship between behavioral problems and classroom justice. On the one hand, behavioral problems led to a decrease in the respectfulness of teacher–child interactions as rated by the teachers. On the other hand, just pedagogical practices as rated by external observers led to decreases in the children's behavioral problems but also in their joy of learning. There was a negative effect of justice sensitivity on joy of learning. These results might be due to teachers' adaptiveness. Teachers might especially focus their attention and efforts on the children who seem to need them the most, for instance, children with a low joy of learning or behavioral problems. These results might be used to sensitize teachers and school developers to the importance of classroom justice.

Keywords Educational justice · Classroom observation · Primary school · Social behavior

✉ Natalie Ehrhardt-Madapathi
  ehrhardt@uni-koblenz.de

[1] Department of Psychology, University of Koblenz-Landau, 56072 Koblenz, Germany

[2] Department of Psychology, University of Koblenz-Landau, Landau, Germany

🍊 Springer

1 Introduction

Justice is a fundamental aspect of any organization because people want to get what they deserve and deserve what they get (Lerner 1977). In Germany, primary school is the first institution that a child has to attend and in which the child has to take up the role of a citizen in a society. The child's role in school is to learn and to integrate into society. Besides the family, the classroom is the child's first social group that he or she has to interact with on a daily basis. Accordingly, how this group organizes itself lays the foundation for the child's sense of justice. In the classroom, rules have to be implemented, and limited resources such as the teacher's attention have to be distributed. The child perceives these interactions and distributions and evaluates them according to their fairness (Dalbert 2013; Tata 1999). Experiencing justice in this first mandatory institution is considered vital for the child to develop trust in society, and it motivates children to behave fairly themselves (Dalbert 2013). Experiences of just or unjust treatment can shape children's worldviews and social behavior, and therefore, schools have a responsibility not only to impart knowledge but also to foster children's social development and to help them integrate into society (Susteck 1996). Accordingly, it is important to achieve classroom justice in primary schools to ensure good developmental conditions. To reach this goal in the long term, classroom justice has to be defined more comprehensively, it has to be measured precisely, and the long-term effects of classroom (in)justice have to be studied. This paper contributes to this goal by presenting the results of a longitudinal study that measured the bidirectional relationship between classroom justice and student outcomes.

1.1 What is classroom justice

Students and teachers agree that classroom justice is important. However, the majority of students claim that they are treated unfairly in school (Dalbert 2011a, b; Israelashvili 1997). This discrepancy between the desire for justice and the experience of injustice in the classroom could be due in part to a lack of a common definition of classroom justice. Many researchers define perceived classroom justice as subjective "perceptions of fairness regarding outcomes or processes that occur in the instructional context" (e.g., Paulsel and Chory-Assad 2005). However, it is reasonable to question whether justice is only subjective. Are there objectively fair or unfair treatments? Whether a given situation is judged as justice-relevant or as just or unjust depends on the dimension and principle of justice that are applied. Therefore, it is important to agree on a common principle of justice that should be applied to justice-relevant situations in the context of a primary school classroom.

1.1.1 Forms of justice

Forms of justice represent contexts in which a situation is justice-relevant. Recent research discriminated between the following four dimensions of justice: distributive, retributive, procedural, and interactional justice (Sabbagh and Schmitt 2016). If

a limited good such as the teacher's time or praise has to be distributed, this is a question of distributive justice (Jasso et al. 2016). Retributive justice is concerned with the punishment of a person who has done something wrong. It defines the appropriate treatment for reinstating justice (Wenzel and Okimoto 2016). Opposed to distributive justice, procedural justice is not related to the results but to the processes that led to a certain outcome (Leventhal 1980; Lind and Tyler 1988; Vermunt and Steensma 2016). Interactional justice deals with the appropriateness and the quality of interactions. Only when the subjects involved are sensitive, kind, and respectful to each other can there be interactional justice (Bies and Moag 1986; Vermunt and Steensma 2016).

Distributive justice is central in the daily routines of school life. Whether a justice-relevant situation is judged as just or unjust depends on the justice principle that is applied to the situation.

1.1.2 Principles of justice

In order to judge whether the outcome of a justice-relevant situation is just or unjust, the forms of justice must be combined with principles. The core principles of distributive justice are equality, equity, and need (Deutsch 1985; Jasso et al. 2016).

The *equality* principle states that everybody should be treated equally and should receive the same outcome (Deutsch 1985). Different from equality, the *equity* principle requires a differentiation (Jasso et al. 2016). If there is equity between the inputs that a person brings to a job and the outcomes that he or she receives from it against the perceived inputs and outcomes of others, then this person will perceive justice. If there is an imbalance that falls toward either side, it can lead to perceived injustice. The *need* principle also requires differentiation (Jasso et al. 2016). According to the need principle, everyone should get what he or she needs in order to succeed and have a decent life.

According to Deutsch (1985), the application of justice principles to justice-relevant situations does not happen randomly. Instead, the principle matches the requirements of the situation. In small and intimate groups, the equality principle is often favored. The equity principle is applied in competitive situations, such as sporting events or in business. The principle of need may be preferred in interactions with people who are unable to acquire their own resources (e.g., sick people, children, the elderly; Gollwitzer and Van Prooijen 2016). In caring-oriented groups, such as a primary school classroom, the need principle is therefore the most appropriate one (Berti et al. 2010; Deutsch 1985; Ehrhardt et al. 2016). A child is treated fairly if his or her needs are satisfied. Typically, the principles of justice are applied to questions of distributive justice only. But in the context of primary school classrooms, it might be appropriate to apply the principle of need also for questions of interactional justice such as the teacher's tone of voice. Children can differ in the extent to which they need a friendly and polite tone. Some children might need the teacher to use a very gentle tone in teacher–child interactions. For other children, a tougher tone might be more appropriate. Since the teacher, other than an employer at an organization, has an explicit educational mandate, he/she has a responsibility to meet the needs of the children.

1.1.3 Aspects of classroom justice

The application of the need principle for distributive justice in the classroom does not yet define which situations can be fair or unfair. According to Resh and Sabbagh (2014, 2016), there are five subspheres of justice: (a) the right to education, (b) educational places (student composition, selection to classes, tracks, ability-based learning groups), (c) pedagogy, (d) grading, and (e) the teacher–student relationship. In this study, the subspheres of justice that act on a microlevel are particularly in focus. In daily classroom routines, these last three subspheres are immediately relevant to the needs of the children. Because no grades are given in the German primary school classroom, we decided to focus on (c) pedagogy and (e) the teacher–student relationship. One of a child's most important needs is to have good learning conditions. Therefore, we focused on just *pedagogical practices* and *teacher–student interactions.* According to Parsons (1959), pedagogical practices are defined as the ways in which the teacher chooses to encourage learning. One aspect of this is to create adaptive learning settings in which the child is offered appropriate tasks and appropriate support. Thorkildsen (1989) attempted to identify classroom practices and asked children to rate the fairness of such practices. All students rated "peer-tutoring" as the fairest pedagogical practice for addressing heterogeneity in the classroom, whereas they reported that "enrichment" was the most frequently applied principle. Resh and Sabbagh (2016) summarized that researchers have yet to study the effects of fair or unfair pedagogical practices. Another one of the child's most important needs is to be accepted and respected by the teacher. This is what we measured with *teacher–student interactions*.

The teacher as the most powerful person in the classroom has a responsibility to ensure that the needs of the children listed above are being met (Sabbagh and Resh 2014). The teacher has to decide how to allocate rewards and punishment. Furthermore, it is one of the teacher's duties to assess students' performance and their behavioral problems. To summarize, classroom justice depends on how well the teacher can apply the principle of need on the level of an individual child (i.e., by applying appropriate pedagogical practices and respectful teacher–child interactions). The teacher, who is in a position of power in the classroom, has a responsibility to ensure classroom justice (Sabbagh and Resh 2014). It is not reasonable to speak of justice in the classroom in general. Rather, justice is manifested or not through individual actions and situations in which a child experiences either just or unjust treatment. Thus, it makes sense to assume that whereas one child's needs may be met in a particular classroom, the needs of another child in the same classroom may be neglected.

1.2 Effects of classroom justice on behavior, well-being, and emotions

What do we know about the effects of classroom justice on children's outcomes? There are not yet many empirical findings on the effects of classroom justice. Dalbert (2011a, b) found that teacher justice had a positive effect on a wide range of outcomes: perceived class climate, perceived social inclusion, academic achievements, and rule-compliant behavior.

Prior research has indicated that classroom justice influences social behavior. Donat et al. (2014) showed that students who reported that they were treated fairly by their teacher were less likely to display disruptive behavior in the classroom and were less likely to cheat on tests. College students who rated classroom procedures as fair were more likely to exhibit a high learning motivation and less aggressive behavior toward the teacher (Chory-Assad 2002). According to a study by Donat et al. (2013), the effect of perceived teacher fairness on students' social behavior is moderated by a feeling of social inclusion and an acceptance of authority.

In addition to children's social behavior, their well-being might also be influenced by classroom justice. Hascher and Edlinger (2009) found that the arrangement of an environment that is conducive to a child's classroom learning—in other words, an adaptive learning setting—is beneficial for the well-being of the students. Peter and Dalbert (2010) found that perceived teacher justice was positively correlated with classroom climate, and classroom climate influenced students' well-being.

Classroom justice might even influence children's joy of learning. Whisenant and Jordan (2008) showed that children who were treated with respect by their physical education teachers reported more joy in doing sports and were more likely to continue to do sports outside of school. According to a study by Molinari et al. (2013), the perceived interactional justice of the teacher had an impact on students' academic achievement as well as on their learning motivation. Primary school children on average experience a great joy of learning. Unfortunately, this joy tends to decrease throughout the school career, and sixth graders already report much less joy of learning than first graders do (Hagenauer 2011; Helmke 1993; Pekrun 1993). A lack of classroom justice could be one of the reasons for this decline in positive emotions toward school and learning.

To conclude, there seem to be three major aspects in children's school life that may be affected by (in)justice. These can be summarized as social behavior, well-being, and joy of learning. Because of these considerations, the present study focused on the effects of (in)justice in school on these student outcomes.

1.3 Justice sensitivity as a possible moderator

1.3.1 Justice sensitivity

Experiencing injustice might be detrimental to students' well-being, social-behavior, and learning motivation. If and how injustice in the classroom is perceived by the children depends on the children's justice sensitivity. Justice sensitivity is a personality trait that affects whether or not a treatment is perceived as unjust, how strong the injustice is considered to be, and the emotional and behavioral reactions to a perceived injustice (Baumert and Schmitt 2016; Schmitt et al. 2005; Van den Bos et al. 2003). People who score high on justice sensitivity are prone to perceiving injustice, and they exhibit greater emotional, cognitive, and behavioral reactions to injustice (Baumert et al. 2011). Justice sensitivity is a personality trait because it is a stable disposition in adults (Schmitt et al. 2010). Nevertheless, little is known about the development of justice sensitivity, its

structure, and its stability in children. In adults (Schmitt et al. 2010) as well as in fifth graders (Pretsch et al. 2015), it was possible to differentiate between four different facets of justice sensitivity that are analogous to the four perspectives of injustice: perpetrator, victim, observer, and beneficiary justice sensitivity. In a classroom, a child can experience injustice from these four different perspectives (Baumert and Schmitt 2016). If a child takes something from a peer, this child experiences injustice from the perspective of a perpetrator. A child who does not receive the teacher's attention when he/she thinks he/she deserves it experiences injustice from the perspective of a victim. The observer perspective comes into play when a child, for example, witnesses another child being wrongfully scolded by the teacher. If a child benefits from an unjust situation, for example, when he/she gets a lot of support from the teacher while the teacher ignores other children's need for support, then the child experiences injustice from the beneficiary perspective.

Attending a classroom in which a child's needs are not being met can lead to a more frequent activation of injustice concepts and—in this way—chronically increase the accessibility of such concepts (cf. Higgins and King 1981). Subsequently, justice sensitivity might increase. Classroom injustice can therefore increase justice sensitivity. However, it is also possible that justice sensitivity might increase the amount and intensity of perceived injustice. Children who perceive more injustice and ruminate for longer periods about injustice are more likely to react to this injustice emotionally and behaviorally.

1.3.2 Justice sensitivity as a moderator in student outcomes

Because justice sensitivity influences how people perceive and react to injustice, it is likely that it will also influence the relationship between classroom justice and student outcomes. Maltese et al. (2013) found that victim sensitive people had a tendency to form the expectation that others have mean intentions and therefore withdrew their cooperation in socially uncertain situations. Justice sensitivity led participants to be more ready to interpret a justice-ambivalent situation as unjust and to react by withdrawing their cooperation. Bondü and Esser (2015) showed that adolescents with ADHD symptoms, thus with behavioral problems, reported significantly higher victim justice sensitivity, lower perpetrator justice sensitivity, and more perceptions of injustice. Schmitt et al. (2008) found that justice sensitivity moderated the effect of the perceived injustice of a job termination and their desire to get revenge on the employer.

Justice sensitive children might be more prone to reacting to injustice with behavioral problems because they experience injustice more intensely and more frequently, and they are more likely to ruminate about their unjust experiences. For students who are high on justice sensitivity, the effects of (in)justice on well-being and joy of learning should be higher (Schmitt and Dörfel 1999). If justice sensitivity were indeed found to moderate the effects of classroom injustice on student outcomes, then teachers might be able to prevent the negative outcomes of inevitable classroom injustice by paying special attention to justice sensitive students.

1.4 Aim of the present study

In line with prior research on educational justice, we expected to find that classroom justice would be found to be relevant to primary school students' behavior, well-being, and joy of learning. To the best of our knowledge, studies have not yet explored the effects of justice on children at the beginning of primary school. Because this period might be especially critical in shaping children's trust in society and their motivation to behave fairly themselves (Dalbert 2013), it is important to further the understanding of the long-term effects of classroom (in)justice on primary school student outcomes in a longitudinal study. There is a lack of causal analytical longitudinal studies on justice in school.

Prior research has relied primarily on single sources of information to measure justice (Correia and Dalbert 2007; Dalbert and Stoeber 2005). It cannot be assumed that primary school children are able to accurately rate classroom justice yet because of their limited conceptual understanding and their difficulties in aggregating and abstracting information from single fair or unfair incidents (Biemer and Lyberg 2003; Piaget 1997). Therefore, we applied a multimethod approach for measuring classroom justice. In addition to using a low inference rating to be provided by observers, we also applied high inference rating instruments for observers and teachers. We measured different aspects of classroom justice from different sources to test which justice ratings were best suited to predict student outcomes. Classroom justice and student outcomes were measured at all three measurement occasions. In correspondence with educational and organizational justice research, we expected that classroom justice would have bidirectional relationships with students' behavioral problems, students' well-being, and their joy of learning. Moreover, we expected that justice sensitivity would moderate the effect of classroom justice on student outcomes. We expected that justice sensitivity would increase the effect of classroom (in)justice on behavioral problems, well-being, and joy of learning.

2 Methods

2.1 Participants

Data was collected from primary students, primary teachers and external observers. Informed written consent was obtained from the principals of the participating schools and from the parents of the participating children. From the children oral consent was obtained. Prior to the data collection the study protocol was approved by the Supervision and Service Administration Body of the states in Germany in which the study was conducted (Rhineland Palantine and Hesse). These institutions approved that the study was of considerable pedagogical and scientific interest, that the burden on the school, the teachers and the students was reasonable, and that the data protection guidelines were met. After the data collection, each class received information about the overall results of the study.

2.1.1 Participating children

A total of 245 students from seven primary schools participated in the study. Altogether, 15 classes took part. Four schools were located in rural areas, whereas the other three were located in urban environments. The children had mixed socioeconomic backgrounds. Convenience sampling was used to find classes whose teachers were willing to participate. In order to take part in the study, the supervising school authority, the school board, the class teacher, the child's parents, and the child him- or herself had to agree to participate. The classes differed remarkably in size: The smallest class consisted of only 14 students, whereas the largest was composed of 30 students. In each class, at least eight and a maximum of 25 children took part in the study. One hundred thirty-two of the students were female (53.8%), the mean age of the students at the first point of measurement was 7.3 years ($SD = .87$ years), and the students attended Grades 1–4. A total of 102 of the students were in a regular first-grade primary school class. The other 143 students attended classes with mixed age groups, in which children from Grades 1–4 were educated together. They attended the following grade levels: Grade 1: $n = 158$, Grade 2: $n = 53$, Grade 3: $n = 8$, and Grade 4: $n = 17$ (missing: $n = 5$).

2.1.2 Participating observers

The observers taking part in the study were either students of psychology or prospective teachers. Altogether, 11 observers participated. Nine of them were female (two male), and their ages ranged from 22 to 47 years ($M = 27.42$, $SD = 6.99$).

2.1.3 Participating teachers

Fifteen primary school teachers (14 female, +++one male) took part in the study. Their mean teaching experience was 9.6 years. All of them were primary school teachers, and all were the classroom teacher in the participating classes.

2.2 Design and data collection

The study had a longitudinal design with three measurement occasions (time lag of 4–5 months each time), which allowed us to examine causal long-term effects of different aspects of classroom justice on children's behavioral problems, well-being, and joy of learning.

For data collection, the first measurement point was at the beginning of the school year in September 2014, the second was in February, and the third was at the end of the school year in June/July during the students' regular school time in the morning.

Depending on the size of the classroom, two to five observers were present in the classroom during the observation. Before the observation started, the observers divided the students into groups, and each observer observed one group of children with up to five children. The observation lasted 120 min, and during the whole

observation time, both the students and the teacher were present in the classroom. The low inference rating (LIR_O) was filled out during the observation period by the external observers. After the observation, the observers filled out the high inference rating instrument (HIR_O). After the school morning, the teachers filled out the high inference rating instrument (HIR_T) and the questionnaire on students' behavioral problems (SDQ-T). On the following day, the children were interviewed and asked to fill out the well-being and joy of learning questionnaires that were read aloud to them.

2.3 Instruments

2.3.1 Justice ratings

In this study, classroom justice was measured with the high and low inference justice ratings from Ehrhardt et al. (2016). The data of the study from Ehrhardt et al. (2016) is the second measurement point of this longitudinal study. These justice ratings measure justice in the classroom on a single-child level. In addition to the high inference justice ratings provided by the teachers (HIR_T) and external raters (HIR_O), the external raters also gave low inference justice ratings (LIR_O) via systematic observations of classroom justice.

Low inference rating systems offer the opportunity to quantify how often a certain category of actions occurs in the classroom. These observable and quantifiable low inference ratings do not require a lot of interpretation from the raters and therefore are often more reliable than high inference rating systems (Lotz et al. 2013). Still, high inference rating systems have some advantages over low inference rating systems. Often, the high inference rating systems can be developed to more closely reflect a particular theory because it is possible to rate the core dimensions of a theoretical construct. This might lead to the high inference ratings having a higher predictive validity (Clausen 2002). In a comparative study by Clausen et al. (2003), the authors found that high inference ratings were better suited to predict student performance.

Because Ehrhardt et al. (2016) considered the need principle to be the appropriate justice principle in a primary school classroom, they developed justice measures that measure whether a child's needs are being met. To accomplish this, they conducted two prestudies, the first with primary school teachers, and the second with first-grade students. In these prestudies, they collected justice-relevant situations and identified situations in which classroom justice could be either observed or rated or both. With factor analysis, they built scales for the rating instruments. For a detailed explanation of the development of the high and low inference justice rating instruments, see Ehrhardt et al. (2016).

Given that there are multiple sources of injustice, a justice instrument should not measure just one overall justice score. Therefore, different aspects of classroom justice were combined into scales according to their factor structure. Doing so made it possible to analyze which aspects of injustice influenced which aspects of the child's development. Three instruments were developed: a low inference justice rating instrument (LIR_O) to be applied by the external observers and high

inference rating instruments to be provided by the teachers (HIR_T) and the external observers (HIR_O). The high inference rating instrument was the same for the teachers and the observers with the exception of wording and phrasing. These were changed in order to present the questions about the teachers' behavior in the first-person singular. The teachers were not explicitly trained to use the high inference rating instrument. They received the instrument before they had to apply it and were given the opportunity to ask any questions they had concerning its use. The teachers had to apply the high inference rating instrument to every child who participated in the study. The teachers were instructed to rate only the justice of the treatment for the morning during which the observation occurred.

2.3.1.1 Observers' and teachers' high inference ratings   Observers and teachers were presented with the questionnaire (high inference justice rating instrument) from Ehrhardt et al. (2016), which contains a total of 12 rating items. The wording of the items that were part of a measurement invariant scale and the name of the scale are given in Table 1. All items were rated on a 4-point rating scale ranging from (1/do not agree at all … 4/completely agree) plus an additional *not applicable* response option. If an item was rated as *not applicable* this rating was coded as missing data. The high inference rating items for observers and teachers were also factor analyzed. In both cases, a three-factor solution was accepted. The teachers' high inference rating factors reflected (a) HIR_T_1 = adaptive learning settings, (b) HIR_T_2 = respectful teacher–child interactions, and (c) HIR_T_3 = ensuring learning opportunities. The observers' high inference rating factors reflected (a) HIR_O_1 = adaptive learning settings, (b) HIR_O_2 = respectful teacher–child interactions, and (c) HIR_O_3 = appropriateness of praise and criticism.

2.3.1.2 Low inference ratings   The low inference rating instrument from Ehrhardt et al. (2016) consists of 17 discrete events and behaviors that are observable in the classroom (e.g., *The child is given the opportunity to speak*). After the observers were trained to use the instrument, their interrater agreement was estimated with the ICC. The interrater agreement was on average .75. This result indicates a satisfactory objectivity of the low inference justice rating instrument. The dimensional structure of the low inference ratings was determined with exploratory factor analysis. Four common factors were extracted, and these were independent of whether or not the class effect was controlled for or not. The extracted factors reflected (1) LIR_O_1 = performance feedback, (2) LIR_O_2 = enforcing class rules, (3) LIR_O_3 = respectful teacher–child interactions, and (4) LIR_O_4 = accepting the child and letting the child act. For a detailed description of the development, the items, and the scales, see Ehrhardt et al. (2016).

2.3.1.3 Measurement invariance   The justice rating instruments are rather new and have not yet been applied in a longitudinal study before. Therefore, it is not yet known whether the scales measure the same construct at every measurement occasion. The scales consist of items that are developed to assess a latent construct. Applied in a longitudinal study, the goal is to follow individuals over time. In order

Table 1 Measurement invariant scales from the high and low inference justice rating instruments

Name of the instrument	Justice rating scale	Items
HIR_T	HIR_T_1 (appropriateness of praise and criticism)	Student is listened to actively
		Praise and criticism are constructive (was deleted)
		Student reacts positively to feedback
		Student gets praise for good performance
HIR_O	HIR_O_1 (adaptive learning settings)	Student is given enough time to think
		Student may finish work that was begun
		Learning arrangements enable individual learning (was deleted)
		Student is given enough opportunities to speak up
	HIR_O_3 (respectful teacher–child interactions)	Teacher's tone is appropriate
		Tone between teacher and student is respectful
		Student is made to finish his or her work (was deleted)
LIR_O	LIR_O_4 (accepting the child and letting the child act)	Feedback concerning social behavior
		Feedback concerning orderliness
		Teacher makes dismissive remark about child (was deleted)
		Opportunity to speak

HIR_T_1 teachers' high inference ratings of the *appropriateness of praise and criticism*, *HIR_O_1* external observers' high inference ratings on the *adaptive learning settings*, *HIR_O_3* external observers' high inference ratings of *respectful teacher–child interactions*, *LIR_O_4* external observers' low inference ratings of the *acceptance of the child*

to be able to validly compare the different measurement occasions, the scales should measure identical constructs with the same structure across different measurement occasions. When this is the case, there is measurement invariance in these measures. Measurement invariance of the instruments is therefore an important requirement for longitudinal studies. As the same instruments were applied at different measurement occasions, measurement invariance had to be tested by using the semTools package .4–6. Multioccasion structural equation models were proposed in order to compare the factor structure of the justice scales and the dependent variables (well-being, behavioral problems, and joy of learning) across the three measurement occasions. The scales without measurement invariance were omitted from further analyses. The fit indices for the invariance tests of the measurement invariant justice scales and the student outcomes are given in the "Appendix" of this article (Table 6). For four of the justice scales, at least weak invariance could be established. In order to ensure measurement invariance, some items had to be removed from the scales. Table 1 gives these four measurement-invariant justice scales with their items, and the items that were removed from the scales are

presented in cursive. HIR_T_1 (adaptive learning settings) was included in further analyses, but the other scales from the high inference ratings provided by teachers were omitted because of a lack of measurement invariance. Out of the high inference ratings provided by external observers, HIR_O_1 and HIR_O_3 turned out to be measurement invariant over time and could be included in further analyses. Out of the scales for low inference justice ratings provided by external observers, only the scale LIR_O_4 showed factorial invariance and could be used in the subsequent cross-lagged panel models. The measurement-invariant scales HIR_T_1 (appropriateness of praise and criticism), HIR_O_3 (respectful teacher–child interactions), and LIR_O_4 (accepting the child and letting the child act) all belonged to the subsphere of educational justice *teacher–child interactions*, and HIR_O_1 (adaptive learning settings) could be categorized as *pedagogical practices.*

2.3.2 Behavioral problems

Behavioral problems (BP) were measured with the German version of Goodman's (1997) Strengths and Difficulties Questionnaire (SDQ-T) completed by the teachers (Klasen et al. 2000). The SDQ-T is a short screening instrument with 25 items that can be matched to five subscales. The subscales are Emotional symptoms (e.g., *often complains about headaches, stomach aches, or sickness*), Conduct problems (e.g., *often fights with other children or bullies them*), Hyperactivity-inattention (e.g., *restless, overactive, cannot stay still for long*), Peer relationship problems (e.g., *rather solitary, tends to play alone*), and Prosocial behavior (e.g., *often offers to help others*) and consist of five items each. The items were scored on a 3-point scale (0 = *not true*, 1 = *somewhat true*, 2 = *certainly true*). Low scores indicate unproblematic social behavior, except for "Prosocial behavior" where higher scores mean more positive social behavior. We did not measure Prosocial behavior because the test authors did not offer the option of including it in the total score because they viewed it as a separate construct. Some items had to be recoded because of their negative phrasing. The subscales Emotional problems, Behavioral problems, Hyperactivity, and Peer relationship problems were grouped together to form the *total difficulties score* that was based on 20 items. The psychometric quality of the SDQ-T turned out to be satisfactory to excellent. The Cronbach's α of the subscales ranged from .72 for Behavioral problems to .90 for Hyperactivity. The internal consistency of the *total difficulties score* ranged from .86 to .87.

2.3.3 Well-being

According to Becker (1994), well-being is always context-specific. A primary school students' well-being at home can be very different from his or her well-being at school. Well-being in school can be measured as a state or a trait construct (Tacke 2006). In studying the long-term effects of classroom justice, the more stable trait of well-being is more suitable than the state of well-being. The trait of well-being consists of the presence of positive emotions and cognitions toward school and the absence of negative emotions and of physical or emotional problems in school

(Hascher and Edlinger 2009). Well-being (WB) in elementary school (children in the first and second grades) was measured with the well-being questionnaire (Wustmann Seiler 2012), which is based on Hascher and Edlinger's (2009) questionnaire. Wustmann's well-being questionnaire consists of a questionnaire for children, a questionnaire for teachers, and a questionnaire for parents (33 items in total). In this study, we used only the questionnaire for children, which consists of 19 items and five subscales: (1) Positive attitudes and emotions toward school (e.g., *I like going to school*), (2) Self-confidence in school (e.g., *To me, anything we do in school is simple*), (3) No worries because of school (e.g., *Lately, I have often worried about how things are going at school*), (4) No social problems in school (e.g., *Recently, other children have hurt me at school*), (5) No physical and psychological complaints (e.g., *How often have you felt sick in the past few weeks?*). The items on the last three subscales had to be recoded to ensure that high scores indicated high well-being in school. In this study, the five subscales were combined into two main scales: *Positive emotions* and *No problems*. The first scale *Positive emotions* contained the subscales Positive attitudes and emotions toward school and Self-confidence in school. The second scale *No problems* contained the following subscales: No worries because of school, No physical and psychological complaints, and No social problems in school. All items were answered on a 4-point scale (for the first four subscales: $1 = strongly\ disagree$ to $4 = strongly\ agree$; for the last subscale: $1 = never$ to $4 = very\ often$). The standardized measurement took place in class where every child was given a questionnaire and had to tick the correct answer. The items and possible answers were read aloud. The psychometric quality of the well-being questionnaire was acceptable to good. The two scales showed internal reliabilities between .76 and .80. The average score across all 19 items had an internal consistency that ranged from .78 to .82.

2.3.4 Joy of learning

The joy of learning in elementary school (JoL) was measured with Rauer and Schuck's (2004) "Fragebogen zur Erfassung emotionaler und sozialer Schuler-fahrungen von Grundschulkindern erster und zweiter Klassen" (Questionnaire to survey the emotional and social experiences of primary school children in first and second grades; FEESS 1-2). The FEESS 1-2 is used to measure the basic emotional and social experiences of schoolchildren in the first and second grades (Rauer and Schuck 2004). It consists of seven subscales; one of them is the Joy of learning subscale (Rauer and Schuck 2003). Joy of learning is defined as experiencing positive emotions while learning. The scale measures joy and positive emotions experienced when doing everyday schoolwork as well as positive attitudes toward schoolwork and the subjects learned in school. Children who score high on the joy of learning subscale have a more positive learning attitude, experience learning as less effortful, and like different types of exercises. High scores mean a high general joy of learning in school. Joy of learning does not refer to specific subjects in school (Rauer and Schuck 2004).

The FEESS 1-2 consists of 90 items in total, and the Joy of learning scale has 13 items (Rauer and Schuck 2004). To measure joy of learning in this study, the following four items were extracted from the scale: *Learning in school is fun*, *I*

enjoy the lessons, *I like my teacher*, and *I like my school*. The items were answered on a 4-point scale (1 = *strongly disagree* to 4 = *strongly agree*). This shortened scale showed internal consistencies between α = .64 and α = .74.

2.3.5 Justice sensitivity

Justice sensitivity is traditionally measured via a questionnaire. There is an eight-item short version (Baumert et al. 2014) that was adapted for children. This instrument was found to be adequate for sixth graders (Pretsch et al. 2015) but proved to be too difficult and abstract for children in primary school. Therefore, we developed a new instrument with vignette stories. Eight justice-relevant situations were combined with each facet of justice sensitivity.

For example, one of these justice-relevant situations was "hiding someone's gym bag." The different facets were: (a) *Imagine that somebody has hidden your gym bag. You have to go to gym class*, *but now you cannot find your bag*. (b) *Imagine that you have hidden another child's gym bag. Now it is time to go to gym, and the other child is searching for his or her bag*. (c) *Imagine that you see two children hiding someone's gym bag. It is time to go to gym class*, *but this child has to search for his or her gym bag now*. (d) *Imagine that some children have hidden the gym bag of the child who was supposed to choose which sports to play today in gym class. Because this child has to search for his/her gym bag*, *he or she cannot attend gym class*, *and therefore*, *you are allowed to choose which sports to play today*. These vignettes were read aloud to each child in a one-on-one interview. After hearing the vignette, the child was asked about the emotional valence of the situation (*How do you feel about this? Good or bad?*), his or her emotional reaction, and the intensity of the perceived emotion. The child was then offered to choose how he or she would like to react. There were four possible options: (1) *help the victim*, (2) *appeal to a judge (teacher or parent)*, (3) *punish the perpetrator*, or (4) *do nothing*. The child's answers were scored as follows: on the emotional level, 2 points for the appropriate emotional valence (bad), 1 point for the appropriate emotional response, up to 3 points for each experienced emotion, up to 1 point for the intensity of the emotion; on the level of the child's behavioral tendency, up to 3 points for the behavioral tendencies. In total, a child could reach 9 points. The postulated four-factor structure of the measurement regarding the four facets of justice sensitivity could not be confirmed by a CFA. Instead, the one-factor solution fit the data best, and we therefore computed a total justice sensitivity score ($\chi^2 = 131.376$, $df = 6$, $p < .001$, RMSEA = .108, CFI = .964, SRMR = .031). Justice sensitivity was assessed at the first measurement occasion.

2.4 Data analysis

Data analysis was conducted with R (R Core Team 2013) and the R packages Lavaan (Rosseel 2012) and semTools (Pornprasertmanit et al. 2014). Cross-lagged panel models were computed in order to investigate the bidirectional effects of the justice rating indices and the outcome variables. Even though the children were not independent units but were instead measured in classes, a prior data analysis showed

no class effects (see Ehrhardt et al. 2016, for more details). Moreover, the number of classes that took part in the study did not result in a sufficient number of Level 2 units to conduct multilevel analyses. Therefore, the multilevel structure was not taken into account in the cross-lagged panel models.

3 Results

3.1 Descriptives

Table 2 presents the descriptives of the measurements at all three measurement occasions. The high inference justice ratings had to be rated on a four-point scale (1/ do not agree at all … 4/completely agree). As depicted, the teachers as well as the external raters rated the level of classroom justice as relatively high. The external observers' low inference ratings for *acceptance of the child* formed a scale that was composed of complex items. Therefore, the scale mean was difficult to interpret. Across the three measurement points, the scores ranged from − 12.00 to .39. A higher number indicates a fairer treatment, and the absolute number could not be interpreted because there was no maximum.

Students' well-being and their joy of learning were high as well, although they showed slight declines over time. Justice sensitivity, on the other hand, increased

Table 2 Descriptives for the measurements at all three measurement occasions

Measure	Scale	T1			T2			T3		
		N	M (SD)	α	N	M (SD)	α	N	M (SD)	α
HIR_T	HIR_T_1	183	3.46 (.63)	.66	150	3.58 (.57)	.56	145	3.52 (.61)	.61
HIR_O	HIR_O_1	206	3.45 (.66)	.66	204	3.47 (.57)	.62	205	3.52 (.54)	.68
	HIR_O_3	207	3.75 (.41)	.66	202	3.87 (1.59)	.58	206	3.79 (.38)	.72
LIR_O	LIR_O_4	207	− .70 (1.30)	.33	207	− .52 (.94)	.33	208	− .47 (.77)	.20
WB	Positive Emotions	216	3.45 (.48)	.77	209	3.42 (.48)	.78	210	3.36 (.49)	.80
	No Problems	216	3.27 (.54)	.77	208	3.19 (.55)	.76	210	3.20 (.56)	.79
	WB total	216	3.35 (.42)	.81	209	3.28 (.42)	.78	210	3.27 (.44)	.82
SDQ	Social behavior	197	5.78 (5.51)	.86	141	5.58 (5.77)	.86	150	5.20 (4.99)	.88
FEESS 1-2	JoL	212	3.53 (.54)	.71	206	3.54 (.50)	.64	212	3.49 (.56)	.74
JS	JS	167	4.72 (.71)	.66	193	5.32 (.85)	.72	206	5.36 (.93)	.76

N number of valid cases, *M* mean, *SD* standard deviation, *T1*, *T2*, and *T3* measurement occasions, *HIR_T_1* teachers' high inference ratings of the *appropriateness of praise and criticism*, *HIR_O_1* external observers' high inference ratings of the *adaptive learning settings*, *HIR_O_3* external observers' high inference ratings of *respectful teacher–child interactions*, *LIR_O_4 1* external observers' low inference ratings of the *acceptance of the child*, *WB* well-being, *JoL* joy of learning, *SDQ* strengths and difficulties questionnaire, *FEESS* 1-2 Questionnaire to survey the emotional and social experiences of primary school children in the first and second grades, *JS* justice sensitivity

over time. The children's social behavioral problems were rated by the teachers on average as low/unremarkable (see Goodman 1997) across all measurement occasions.

3.2 Correspondence between the justice ratings and the outcome variables

Table 3 presents the correlations of all measures at the three measurement occasions. There were small, positive correlations between the different justice measures at all three measurement points. The correlations of the same justice measures at different measurement occasions were higher ($0 < r < .527^{**}$) than the ones between the different justice ratings that were measured at the same time ($0 < r < .328^{**}$). The teachers' high inference ratings did not show significant correlations with the observers' high inference ratings. Behavioral problems were negatively correlated with the justice measures ($-.506^{**} < r < 0$), whereas the small correlations for the justice ratings with well-being ($0 < r < .239^{**}$) and joy of learning ($0 < r < .202^{**}$) imply that well-being and joy of learning are significantly related to justice but not strongly. This could be due to the fact that well-being and joy of learning were rated by the students and the behavioral problems were rated by the teachers, just as the HIR_T_1 (adaptive learning settings) were.

3.3 Cross-lagged panel models for the association between justice ratings and student outcomes including justice sensitivity as a covariate

Using latent variables, cross-lagged panel models were specified to investigate the bidirectional effects of the justice rating indices and the outcome variables. Relationships between the justice ratings and student outcomes were estimated. Because there were no theoretical assumptions that suggested different effects across the different measurement occasions, equality constraints were imposed across the three different measurement occasions. The same procedure was conducted for all models. The loadings for the repeated indicators were set equal over time, and the correlated measurement residuals among the repeated indicators were estimated to account for the correlated uniqueness of the indicators over time.

Table 4 presents a summary of the cross-lagged effects of the justice rating indices and the outcome variables. Three models showed a good fit and significant effects: BP_HIR_T_1, BP_HIR_O_1, and JoL_HIR_O_3. The other models either had no cross-lag effect, fit poorly, or had a lack of convergence, which was due to two or more variables that were so highly correlated that the model could not be estimated. The models with an acceptable fit are presented in Figs. 1, 2 and 3.

3.3.1 Bidirectional effects between BP and HIR_T_1 (appropriateness of praise and criticism)

Figure 1 presents the cross-lagged panel model for the association between HIR_T_1 and behavioral problems. In the interest of clarity, the error terms and error correlations are not presented. The model Chi square was significant, but the descriptive fit indices suggested an acceptable fit. The autoregressive paths were

Table 3 Spearman correlations between the justice measures, the student outcomes (joy of learning, well-being, and behavioral problems) and justice sensitivity at all three measurement occasions

Measures	02	03	04	05	06	07	08	09	10	11	12
01 T1_HIR_T_1	.40**	.30**	− .08	− .22**	− .06	.04	.01	.00	.10	.08	− .01
02 T2_HIR_T_1	1.00	.21*	.19*	− .02	.06	.12	− .03	− .02	.07	.14	.11
03 T3_HIR_T_1		1.00	.06	.00	− .02	.05	.07	− .03	− .04	− .07	− .10
04 T1_HIR_O_1			1.00	.18*	− .02	.32**	.09	.09	.11	.21**	.12
05 T2_HIR_O_1				1.00	.15*	.02	.08	.17*	.03	.04	.10
06 T3_HIR_O_1					1.00	− .03	.10	.18*	− .09	.06	.03
07 T1_HIR_O_3						1.00	− .11	.07	.17*	.13	.11
08 T2_HIR_O_3							1.00	.13	.14	− .09	.04
09 T3_HIR_O_3								1.00	.30**	.41**	.33**
10 T1_LIR_O_4									1.00	.30**	.38**
11 T2_LIR_O_4										1.00	.53**
12 T3_LIR_O_4											1.00
13 T1_BP											
14 T2_BP											
15 T3_BP											
16 T1_WB											
17 T2_WB											
18 T3_WB											
19 T1_JoL											
20 T2_JoL											
21 T3_JoL											
22 JS											

Table 3 continued

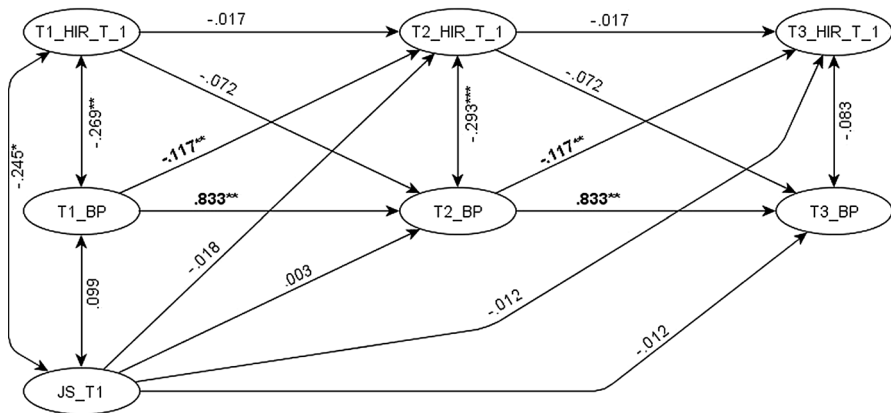|	Measures									
	13	14	15	16	17	18	19	20	21	22
01 T1_HIR_T_1	−.27**	−.18*	−.12	−.04	−.02	−.09	−.02	.02	−.05	.25**
02 T2_HIR_T_1	−.24**	−.29**	−.18	−.03	−.05	−.04	.20*	.01	−.04	.26**
03 T3_HIR_T_1	−.11	−.03	−.08	−.10	−.01	−.16	−.05	−.07	−.02	.11
04 T1_HIR_O_1	−.02	.00	−.12	−.12	−.13	.01	.03	.05	.00	.08
05 T2_HIR_O_1	−.08	−.07	−.01	.08	.13	.00	.02	.00	.08	.02
06 T3_HIR_O_1	−.16*	−.25**	−.15	.02	.03	.07	.09	−.12	.08	.03
07 T1_HIR_O_3	−.12	−.23*	−.09	.01	.00	.01	−.06	.01	−.07	.01
08 T2_HIR_O_3	−.10	.09	.05	.02	.11	−.05	.01	.02	−.12	.09
09 T3_HIR_O_3	−.16*	−.21*	−.17*	−.03	.00	.02	−.11	.12	.14*	.18*
10 T1_LIR_O_4	−.25**	−.22*	−.21*	−.05	.14	.03	.07	.12	.11	.12
11 T2_LIR_O_4	−.22**	−.32**	−.28**	−.08	−.03	.14*	.01	.01	.12	.20*
12 T3_LIR_O_4	−.29**	−.51**	−.38**	.01	.02	.24**	.05	.20**	.09	.09
13 T1_BP	1.00	.81**	.69**	−.18*	−.25**	−.33**	−.16*	−.20**	−.22**	−.18*
14 T2_BP		1.00	.81**	−.25**	−.30**	−.41**	−.19*	−.15	−.17	−.18
15 T3_BP			1.00	−.04	−.27**	−.32**	−.10	−.14	−.15	−.09
16 T1_WB				1.00	.40**	.24**	.32**	.18*	.12	.07
17 T2_WB					1.00	.32**	.17*	.29**	.31**	.02
18 T3_WB						1.00	.20**	.24**	.45**	.05
19 T1_JoL							1.00	.32**	.27**	.11
20 T2_JoL								1.00	.36**	.06
21 T3_JoL									1.00	.17*
22 JS										1.00

T1, *T2*, and *T3* measurement occasions, *HIR_T_1* teachers' high inference ratings of the *appropriateness of praise and criticism*, *HIR_O_1* external observers' high inference ratings of the *adaptive learning settings*, *HIR_O_3* external observers' high inference ratings of *respectful teacher–child interactions*, *LIR_O_4 1* external observers' low inference ratings of the *acceptance of the child*, *BP* behavioral problems, *WB* well-being, *JoL* joy of learning, *JS* justice sensitivity

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

Table 4 Summary of the cross-lagged effects of justice measurements and behavioral problems, well-being, and joy of learning

| SO	Justice ratings	N	CFI	RMSEA	SRMR	Effect
BP	HIR_T_1	246	.896	.056	.087	Small effect of BP on HIR_T_1
	HIR_O_1	246	.929	.045	.086	Effect of HIR_O_1 on BP
	HIR_O_3	242	.966	.036	.101	Effect of HIR_O_3 on BP
	LIR_O_4	242	Model did not converge			Effect of LIR_O_4 on BP
WB	HIR_T_1	242	.899	.047	.080	Effect of WB on HIR_T_1
	HIR_O_1	241	Model did not converge			
	HIR_O_3	241	1.000	.000	.065	Effect of WB on HIR_O_1
	LIR_O_4	241	.626	.178	.131	No effect
JoL	HIR_T_1	246	.854	.049	.087	No effect
	HIR_O_1	241	.905	.039	.075	No effect
	HIR_O_3	246	.935	.035	.078	Small negative effect of JoL on HIR_O_3
	LIR_O_4	246	.450	.173	.121	Small effect of JoL on LIR_O_4

SO student outcomes, *N* number of observations, *CFI* Comparative Fit Index, *RMSEA* Root Mean Square Error of Approximation, *SRMR* Standardized Root Mean Square Residual, *BP* behavioral problems, *WB* well-being, *JoL* joy of learning, *HIR_T_1* teachers' high inference ratings of the *appropriateness of praise and criticism*, *HIR_O_1* external observers' high inference ratings of *adaptive learning settings*, *HIR_O_3* external observers' high inference ratings of *respectful teacher–child interactions*, *LIR_O_4* external observers' low inference ratings of the *acceptance of the child*

Fig. 1 Cross-lagged panel model for HIR_T_1 and behavioral problems with JS as a covariate. $\chi^2 = 334.066$ ($N = 246$, missing patterns = 85), $p < .001$, CFI = .896, SRMR = .087, RMSEA = .056. T1, T2, and T3, measurement occasions; HIR_T_1, teachers' high inference ratings of the appropriateness of praise and criticism; BP, behavioral problems; JS, justice sensitivity. $^\dagger p < .10$; $*p < .05$; $**p < .01$
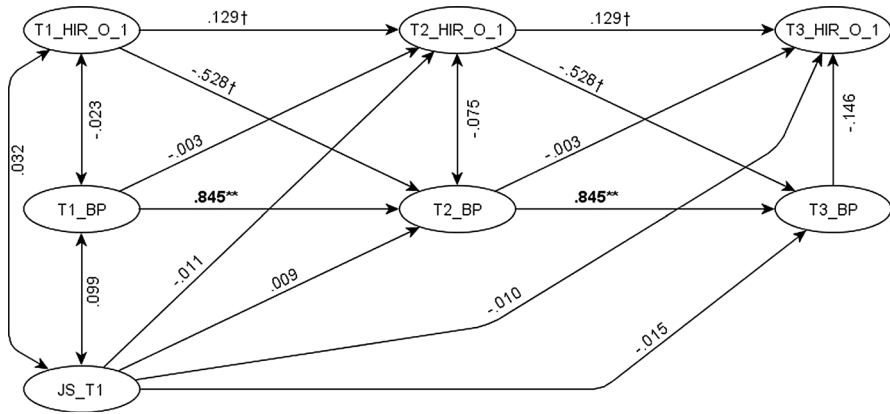
Fig. 2 Cross-lagged panel model for HIR_O_1 and behavioral problems with JS as a covariate. $\chi^2 = 279.836$ ($N = 246$, missing patterns = 92), $p < .001$, CFI = .929, SRMR = .086, RMSEA = .045. T1, T2, and T3, measurement occasions; HIR_O_1, external observers' high inference ratings of the adaptive learning settings; BP, behavioral problems; JS, justice sensitivity. [†]$p < .10$; *$p < .05$; **$p < .01$
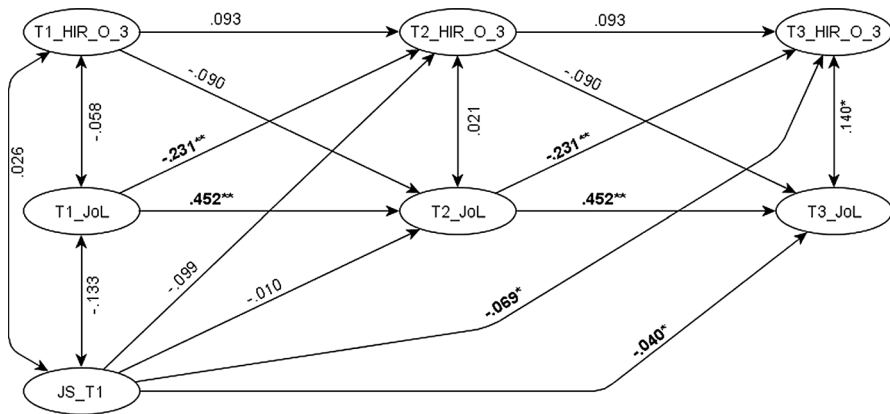
Fig. 3 Cross-lagged panel model for HIR_O_3 and joy of learning with JS as a covariate. $\chi^2 = 244.800$ ($N = 246$, missing patterns = 59), $p < .001$, CFI = .935, SRMR = .078, RMSEA = .035. T1, T2, and T3, measurement occasions; HIR_O_3, external observers' high inference ratings of respectful teacher–child interactions; JoL, joy of learning; JS, justice sensitivity. [†]$p < .10$; *$p < .05$; **$p < .01$

significant for behavioral problems and for HIR_T_1. The total variance accounted for in HIR_T_1 was approximately 12%, and in behavioral problems, it was approximately 82% at Measurement Occasion 2. For Measurement Occasion 3, the total variance accounted for was 16% in HIR_T_1 and 77% in behavioral problems. Behavioral problems were stable over time, whereas there was no continuity in the teachers' justice ratings. Behavioral problems had a negative effect on HIR_T_1. The teachers themselves reported that their praise and criticism were less appropriate for children with behavioral problems.

3.3.2 Bidirectional effects between BP and HIR_O_1 (adaptive learning settings)

Figure 2 presents the cross-lagged panel model for the association between HIR_O_1 and behavioral problems. The model Chi square was significant, but the descriptive fit indices suggested a good fit. The autoregressive paths were significant for behavioral problems and for HIR_O_1. The total variance accounted for in HIR_O_1 was approximately 4%, and in behavioral problems, it was approximately 82% at Measurement Occasion 2. For Measurement Occasion 3, it was 4% for HIR_O_1 and 76% for behavioral problems.

Behavioral problems were stable over time, whereas there was no continuity in the adaptiveness of the learning settings. There was a negative effect of HIR_O_1 on behavioral problems. The students who were treated fairly exhibited fewer behavioral problems at later measurement points.

3.3.3 Bidirectional effects between JoL and HIR_O_3 (respectful teacher–child interactions)

Figure 3 presents the cross-lagged panel model for the association between HIR_O_3 and student's joy of learning. The Chi square was significant, but the descriptive fit indices suggested a good fit. The autoregressive paths were significant for joy of learning but not for HIR_O_3. The total variance accounted for in HIR_O_3 was approximately 23%, and in joy of learning, it was approximately 31% at Measurement Occasion 2. For Measurement Occasion 3, it was 9% for HIR_O_3 and 28% for joy of learning. There was a negative effect of joy of learning on teacher–child interactions. The children who exhibited a high joy of learning were treated less fairly by their teacher. Justice sensitivity had small negative effects on HIR_O_3 as well as on joy of learning.

3.4 Effects of justice sensitivity

To test whether or not justice sensitivity would be found to moderate the relationship between the justice ratings and student outcomes, we included justice sensitivity in the model and tested for main effects and interactions. Contrary to our hypotheses, justice sensitivity did not moderate the relationship between the justice ratings and students' outcomes. Instead, there was a main effect of justice sensitivity in the following models: JoL_HIR_T_1, JoL_HIR_O_1, and JoL_HIR_O_3. The corresponding models are presented in Table 5.

Justice sensitivity had a negative effect on joy of learning in all three models ($-.031$**, $-.033$**, $-.040$**) and also negatively affected HIR_O_3 (*respectful teacher–child interactions*; $-.069$**). No main effect or moderating effect of justice sensitivity on behavioral problems or well-being was found.

Table 5 summary of the effects of justice sensitivity on the justice measurements and BP, WB, and JoL

SO	Justice ratings	N	CFI	RMSEA	SRMR	Effect
JoL	HIR_T_1	247	.854	.049	.087	Small negative effect (− .031**) of JS on JoL
	HIR_O_1	248	.905	.039	.075	Small negative effect (− .033**) of JS on JoL
	HIR_O_3	241	.924	.045	.081	Small negative effect (− .040**) of JS on JoL, and small negative effect (− .069**) of JS on HIR_O_3

SO student outcomes, *N* number of observations, *CFI* Comparative Fit Index, *RMSEA* Root Mean Square Error of Approximation, *SRMR* Standardized Root Mean Square Residual, *JoL* joy of learning, *HIR_T_1* teachers' high inference ratings of the *appropriateness of praise and criticism*, *HIR_O_1* external observers' high inference ratings of *adaptive learning settings*, *HIR_O_3* external observers' high inference ratings of *respectful teacher–child interactions*

4 Discussion

To our knowledge, this is the first study to examine the long-term bidirectional relationship between classroom justice and student outcomes in a sample of primary school students. The results can be regarded as the first evidence that there is a bidirectional relationship between classroom justice and students' behavioral problems. Just treatment decreased behavioral problems, whereas behavioral problems decreased the fairness of the treatment. The results also provide a first indication that there is a negative effect of classroom justice on students' joy of learning. As classroom justice increases, joy of learning decreases. These results support the hypothesis that classroom justice is vital for primary school students' development.

In addition, this study examined the moderating effects of justice sensitivity on the relationship between classroom justice and student outcomes. In contrast to research on adolescent students and adults, no interaction effect was found. We will first discuss the justice rating instruments and then the long-term effects of classroom justice on student behavior and their joy of learning.

4.1 Justice measures

We applied low and high inference ratings of classroom justice across three measurement occasions over the course of a school year and had different judges perform the ratings (observers and teachers). The different measures of classroom justice had a limited overlap (see Table 3 and Ehrhardt et al. 2016, for more details). This implies that different justice ratings provide unique and distinct information and that one measure alone cannot capture classroom justice comprehensively. Future research should therefore include multiple perspectives and methods of measuring classroom justice because one perspective can possibly have important effects on students' outcomes that are not captured by another. Accordingly, our results showed that high inference justice ratings provided by teachers revealed cross-lagged effects different from the high inference justice ratings provided by

external observers. As presented in Fig. 3, the negative effect of joy of learning on the high inference ratings of the *teacher–child interactions* was found for the ratings provided by the observers but not the teachers. The low inference justice ratings revealed no connection to student outcomes. Therefore, the low inference justice rating instrument needs further improvement to increase its predictive validity. It might had been beneficial to include more items in the low inference rating instrument and to make these items more specific. We expect that there are benefits from attempting to measure classroom justice in an objective manner with low inference ratings. Therefore, the low inference rating instrument should be revised to improve its reliability and predictive validity.

The autoregressive effects of the justice ratings were low. This indicates a low temporal stability of classroom justice. Thus, it can be assumed that classroom justice is not stable over time and that the degree of just treatment that a child perceives on 1 day is almost independent of the degree of just treatment on the next day. However, we found significant cross-lagged effects with time lags of 4–5 months. Even though the autoregressive effects were low, measurement invariance was established for four justice rating scales (HIR_O_1 *adaptive learning settings*, HIR_O_3 *teacher–child interactions*, HIR_T_1 *appropriateness of praise and criticism*, and LIR_O_4 *acceptance of the child*), and cross-lagged effects of classroom justice were found. This is indicative of the high quality of the high inference ratings that were used.

4.2 Bidirectional relationship between classroom justice and behavioral problems, well-being, and joy of learning

4.2.1 Relationship between classroom justice and behavioral problems

We found a bidirectional relationship between classroom justice and students' behavioral problems. Children reacted to unfair treatment concerning adaptive learning settings with behavioral problems, and vice versa. Behavioral problems led to a decrease in classroom justice concerning the appropriateness of praise and criticism, thus creating a vicious circle. A recent study showed very similar results: Children who received more positive than negative feedback from their teacher exhibited an increase in prosocial behavior during the course of a school year. Children who received more negative feedback in turn exhibited a greater increase in disruptive behavior (Reinke et al. 2016). This result is in keeping with research on social and emotional development in early childhood. Results in this area show a bidirectional relationship between children and their environment: Children's development is affected by their environment, but children also shape their environment by evoking certain reactions in the people around them (Benson and Haith 2009).

An adaptive learning setting could decrease children's behavioral problems. This is in line with findings from organizational justice research: Experienced injustice often led to counterproductive work behavior in employees (Furnham and Siegel 2012; Greenberg 1990), whereas high organizational justice promoted organizational citizenship behavior (Folger and Cropanzano 1998; Greenberg and Colquitt

2005). When a child experiences unfair treatment, and if this threatens the child's well-being and makes the child feel misunderstood, betrayed, or humiliated, it is a natural reaction for a child to act out (Shaw et al. 2003). Behavioral problems can therefore offer a way to react to unfair circumstances that produce negative emotions. Correspondingly, Skalická et al. (2015) found reciprocal effects in a longitudinal study that linked child-teacher conflict and behavioral problems.

On the other hand, in the current study, behavioral problems led to decreased interactional justice according to the high inference teacher ratings. Similarly, Rudasill et al. (2006) found that a child's temperament could predict the teacher–child relationship. The child's behavior evoked a response in the teacher who had to interact with the child. This effect was found for the high inference ratings provided by the teachers but not for the high inference ratings provided by the external observers. Therefore, the teacher felt that he or she did not treat the children with behavioral problems in an adequate and just way, but apparently, this did not show up in the observer ratings. In this case, it is likely that teachers are more sensitive to their behavior toward students with behavioral problems than external observers are.

4.2.2 Relationship between classroom justice and students' well-being

We found no significant relationship between any justice rating scale and students' well-being. Whereas students' behavioral problems were assessed with a questionnaire filled out by the teachers, well-being was assessed by the children themselves. This could be a methodological reason for why no relationship between the high inference justice ratings and the children's well-being was found. In addition, primary school children might not yet be able to accurately fill out a questionnaire on well-being because they struggle with aggregating and abstracting information from single incidents (Biemer and Lyberg 2003; Piaget 1997). Another possible explanation is that a child's well-being is determined by multiple factors, which makes it unlikely that a single factor (e.g., classroom justice) will have a very strong effect. Because these children were still at the beginning of primary school, their well-being was still generally high, and the children might not yet have experienced a lot of injustice in the classroom that could have changed their general well-being. The main reason for the lack of correlation might be that there is a ceiling effect for the students' well-being ratings. Ceiling effects decrease variability and therefore make relationships between constructs harder to detect.

4.2.3 Relationship between classroom justice and joy of learning

We found a negative effect of joy of learning on the high inference ratings provided by observers concerning the measurement of respectful teacher–child interactions. Children who exhibited a high joy of learning were treated with less respect by their teachers. Behrensen (2013) studied how teachers distribute the precious and rare good of teacher attention, and his arguments are in line with what we found in this study: Through a policy change in Germany and a change in the composition of primary school classrooms in recent years, teachers have to

cope with increasingly heterogeneous groups of students. Teachers are expected to teach in an inclusive manner. Although this is an opportunity to increase the social justice of access to education, it can pose an additional challenge to the teacher in the classroom. He or she has to divide his/her attention carefully and cannot simply use a "one-treatment-fits-all" policy. It could be the case that the teacher is adaptive in his or her behavior and divides attention according to the need principle. He or she might focus his/her attention and efforts on children in need, for instance, the children with a low joy of learning. A child who apparently experiences a high joy of learning might not be in need of the teacher's attention and might therefore perceive less respectful teacher–child interactions. Effects of joy of learning on justice ratings occurred only for the high inference ratings provided by the observers but not for the ones provided by teachers. Observers provided lower justice ratings for the children with a high joy of learning and teachers provided lower justice ratings for the children with behavioral problems. Albeit, all in all, the teachers were not more critical of themselves than the external observers were of the teachers. This can be seen in the means of the high inference justice ratings, which were equally high when provided by the teachers as when provided by the external observers.

4.3 Justice sensitivity and its effects

A new instrument was developed to measure justice sensitivity in primary school children. In contrast to existing measures for adults and adolescents (Baumert et al. 2014; Pretsch et al. 2015), we could not establish a four-factor structure for the four facets of justice sensitivity. There are two possible reasons for this. Either the instrument is not well suited to discriminate between the different facets of justice sensitivity, or primary school children do not yet have a distinct understanding of the different facets and therefore cannot yet discriminate between them consistently. Further research is needed, to answer this question. Most likely, this relationship should be assessed in an older age group because, in older children, the different facets of justice sensitivity are more likely to already be manifest. Because our measures were not able to discriminate between the different facets of justice sensitivity, we could not answer the question of whether different facets affect the relationship between classroom justice and student outcomes in distinct ways. Instead, we tested whether the overall level of justice sensitivity had a moderating effect on the relationship between classroom justice and student outcomes. We expected that justice sensitivity would moderate the effects of classroom justice on student outcomes. This moderating effect was not found. However, we found a negative main effect of justice sensitivity on children's joy of learning and on the measure of *respectful teacher–child interactions*. Children who are justice sensitive are more prone to perceiving injustice, and they experience stronger emotional, cognitive, and behavioral reactions to injustice. Justice sensitivity is an important predictor of protest behavior against injustice and inequality (Baumert and Schmitt 2009; Baumert et al. 2011). Thus, this result is not surprising. A child who is busy ruminating about an experience of injustice is less likely to experience joy of

learning at the same time. More research is needed to affirm the negative effect of justice sensitivity on students' joy of learning.

4.4 Limitations of the study

To our knowledge, this study was the first to measure classroom justice with a multimethod approach, measuring justice from different perspectives. However, we faced some methodological problems. One problem is the ceiling effect of the high inference justice ratings. Although a high justice rating is generally good for the children, it makes it more difficult to detect effects of classroom justice on student outcomes. According to Terwee et al. (2007), there is a ceiling effect if more than 15% of respondents achieve the highest possible score. This was the case for the high inference justice ratings in our study. The high inference justice ratings were made on a scale that ranged from 1 to 4, and the mean high inference ratings were between 3.45 and 3.87. The external raters as well as the teachers themselves rated the degree of just treatment as relatively high for all students. The presence of a ceiling effect indicates that extreme items at the upper end of the scale are missing. This may limit the content validity. Consequently, it was not possible to discriminate between different degrees of fair treatment, and thus, reliability was reduced. The rating instruments should therefore be adapted before using them in further studies. One option for addressing the ceiling effect would be to apply a larger Likert scale (e.g., a 6-point Likert scale). More answer options might increase the variability of the scales. However, an effect of classroom justice on student outcomes was still established in our data.

Another limitation of the study is its data structure. Even though the children were nested in classes, there were not enough classes in the study to conduct a multilevel analysis. It would have been interesting to discriminate between the within-class and between-class effects of classroom justice, but even though there was no multilevel analysis, prior analyses of the data showed no class effects (Ehrhardt et al. 2016).

4.5 Conclusions and practical implications

Future research should focus on the moderating role of justice sensitivity on the effects of classroom justice. The different facets of justice sensitivity might have specific effects on student outcomes. Highly victim sensitive students are more likely to react to classroom injustice with retaliatory behavior, whereas beneficiary sensitive students might react to injustice by offering to help the victims in order to restore justice.

Further studies could apply experimental conditions in order to assess the effects of classroom justice under more controlled settings and in order to test for moderating effects in more detail. Classroom injustice is an unavoidable part of students' everyday lives. It is impossible to ensure that every child's needs are met every single time. Nevertheless, fortunately, all of the justice ratings in the current study were relatively high.

The results of this study provide a first indication that injustice in the classroom has a negative effect on students' social behavior. Especially the bidirectional relationship of behavioral problems and classroom injustice is problematic because it might lead to a vicious circle. These results should be used to sensitize teachers and school developers to the importance of classroom justice. To ensure good learning conditions for students, teachers should be aware of the importance of classroom justice, they should be able to assess classroom justice, and they should be motivated and capable of restoring justice in the classroom. Teachers should also be aware of the multiple perspectives of classroom justice. The amount of overlap in the high inference justice ratings provided by the teachers and by the external observers was low. This implies that even a teacher who is aware of the importance of classroom justice and who assesses his/her own behavior toward the children as just might still be considered unjust by external observers or by the students themselves. One way to enable teachers to enforce classroom justice could be to add the topic of classroom justice to the teacher–training curriculum.

Appendix: Measurement invariance test

Table 6 provides the fit indices for the invariance tests for all models. Model 1 measured configural invariance, Model 2 measured weak invariance (equal loadings), Model 3 measured strong invariance (equal loadings and intercepts), Model 4 measured strict invariance (equal loadings + intercepts + residuals), and Model 5 measured full factor mean invariance (equal loadings + intercepts + residuals + means). According to Kline (2015), the fit of the measurement invariance model is adequate if the CFI is > .90 and the cut-off value for the RMSEA is < .08. When the model fit was assessed according to these rules, the following scales showed at least configural invariance: HIR_T_1, HIR_O_1, HIR_O_3, and LIR_O_4.

Table 6 Measurement invariance tests for the models including the justice ratings and the student outcomes

Scales	Model	Chi² (df)	p value	CFI	RMSEA	Comparison	Decision
BP_HIR_T_1	1	45.934 (24)	.005	.953	.086		Accept
	2	62.007 (32)	.001	.938	.086	Model 1 versus 2	Reject
BP_HIR_O_1	1	22.492 (24)	.550	1.000	.000		Accept
	2	29.256 (32)	.606	1.000	.000	Model 1 versus 2	Accept
	3	35.058 (40)	.692	1.000	.000	Model 2 versus 3	Accept
	4	81.851 (52)	.005	.924	.067	Model 3 versus 4	Reject
BP_HIR_O_3	1	20.492 (12)	.058	.983	.070		Accept
	2	24.966 (18)	.126	.986	.052	Model 1 versus 2	Accept
	3	28.508 (24)	.239	.991	.036	Model 2 versus 3	Accept
	4	56.762 (34)	.008	.956	.068	Model 3 versus 4	Reject
BP_LIR_O_4	1	64.443 (24)	.000	.913	.107		Reject
WB_HIR_T_1	1	17.208 (12)	.142	.963	.059		Accept
	2	21.248 (18)	.267	.977	.038	Model 1 versus 2	Accept
	3 (model did not converge)						Reject
WB_HIR_O_1	1	19.794 (12)	.071	.956	.062		Accept
	2 (model did not converge)						Reject
WB_HIR_O_3	1 (model did not converge)						Reject
WB_LIR_4	1	64.443 (24)	.0000	.913	.107		Accept
	2	79.232 (32)	.000	898	.100	Model 1 versus 2	Accept
	3	86.865 (40)	.000	.899	.089	Model 2 versus 3	Accept
	4	207.885 (52)	.000	.664	.143	Model 3 versus 4	Reject
JoL_HIR_T_1	1	48.198 (39)	.148	.980	.042		Accept
	2	68.502 (49)	.034	.958	.055	Model 1 versus 2	Reject

Table 6 continued

Scales	Model	Chi² (df)	p value	CFI	RMSEA	Comparison	Decision
JoL_HIR_O_1	1	71.933 (39)	.001	.945	.069		Accept
	2	98.044 (49)	.000	.918	.075	Model 1 versus 2	Reject
JoL_HIR_O_3	1 (model did not converge)						Reject
JoL_LIR_4	1	38.930 (39)	.473	1.000	.001		Accept
	2 (model did not converge)						Reject

df degrees of freedom, *CFI* Comparative Fit Index, *RMSEA* Root Mean Square Error of Approximation, *SRMR* Standardized Root Mean Square Residual, *BP* behavioral problems, *WB* well-being, *JoL* joy of learning, *HIR_T_1* teachers' high inference ratings of the *appropriateness of praise and criticism*, *HIR_O_1* external observers' high inference ratings of *adaptive learning settings*, *HIR_O_3* external observers' high inference ratings of *respectful teacher–child interactions*, *LIR_O_4* external observers' low inference ratings of the *acceptance of the child*

References

Baumert, A., Beierlein, C., Schmitt, M., Kemper, C. J., Kovaleva, A., Liebig, S., et al. (2014). Measuring four perspectives of justice sensitivity with two items each. *Journal of Personality Assessment, 96*(3), 380–390. https://doi.org/10.1080/00223891.2013.836526.

Baumert, A., Gollwitzer, M., Staubach, M., & Schmitt, M. (2011). Justice sensitivity and the processing of justice-related information. *European Journal of Personality, 25*(5), 386–397. https://doi.org/10.1002/per.800.

Baumert, A., & Schmitt, M. (2009). Justice-sensitive interpretations of ambiguous situations. *Australian Journal of Psychology, 61*(1), 6–12. https://doi.org/10.1080/00049530802607597.

Baumert, A., & Schmitt, M. (2016). Justice sensitivity. In C. Sabbagh & M. Schmitt (Eds.), *Handbook of social justice theory and research* (pp. 161–180). New York, NY: Springer. https://doi.org/10.1007/978-1-4939-3216-0_9.

Becker, P. (1994). Theoretische Grundlagen [Theoretical basics]. In A. Abele & P. Becker (Eds.), *Wohlbefinden* (pp. 13–49). Weinheim: Juventa.

Behrensen, B. (2013). Die Verteilung der Lehreraufmerksamkeit als Gerechtigkeitsfrage –Sekundäranalyse einer Studie zu Sichtweisen und Erfahrungen von Grundschullehrkräften mit individueller Förderung [The allocation of teacher attention as a question of justice—Secondary analysis of a study on attitudes and experiences of primary school teachers on the individual support of students]. *Pädagogische Rundschau, 67*(5), 553–562.

Benson, J. B., & Haith, M. M. (Eds.). (2009). *Social and emotional development in infancy and early childhood* (1st ed.). Amsterdam: Elsevier.

Berti, C., Molinari, L., & Speltini, G. (2010). Classroom justice and psychological engagement: Students' and teachers' representations. *Social Psychology of Education, 13*(4), 541–556.

Biemer, P. P., & Lyberg, L. E. (2003). *Introduction to survey quality*. New Jersey: Wiley.

Bies, R. J., & Moag, J. S. (1986). Interactional justice: Communication criteria of fairness. In R. J. Lewicki, B. H. Sheppard, & M. Bazerman (Eds.), *Research on negotiation in organizations* (pp. 43–55). Greenwich, CT: JAI Press.

Bondü, R., & Esser, G. (2015). Justice and rejection sensitivity in children and adolescents with ADHD symptoms. *European Child and Adolescent Psychiatry, 24*(2), 185–198. https://doi.org/10.1007/s00787-014-0560-9.

Chory-Assad, R. M. (2002). Classroom justice: Perceptions of fairness as a predictor of student motivation, learning, and aggression. *Communication Quarterly, 50*(1), 58–77.

Clausen, M. (2002). *Unterrichtsqualität: Eine Frage der Perspektive? Empirische Analysen zur Übereinstimmung, Konstrukt- und Kriteriumsvalidität [Instructional quality: A question of one's perspective? Empirical analyses on concurrent, construct and criterion validity]*. Münster: Waxmann.

Clausen, M., Reusser, K., & Klieme, E. (2003). Unterrichtsqualität auf der Basis hoch-inferenter Unterrichtsbeurteilungen: Ein Vergleich zwischen Deutschland und der deutschsprachigen Schweiz [Instructional quality based on high inference ratings: A comparison of Germany and German-speaking Switzerland]. *Unterrichtswissenschaft, 31*(2), 122–141.

Correia, I., & Dalbert, C. (2007). Belief in a just world, justice concerns, and well-being at Portuguese schools. *European Journal of Psychology of Education, 22*(4), 421–437.

Dalbert, C. (Eds.) (2011). Über die Bedeutung des schulischen Gerechtigkeitserlebens [On the significance of experienced justice in school]. In *Sozialpsychologie, Sozialisation und Schule: Beiträge des 26. Hamburger Symposions zur Methodologie der Sozialpsychologie* (pp. 217–232). Lengerich: Pabst Science Publisher.

Dalbert, C. (2011b). Warum die durch die Schüler und Schülerinnen individuell und subjektiv erlebte Gerechtigkeit des Lehrerhandelns wichtig ist [Why subjectively perceived teacher justice is important]. *Zeitschrift für Pädagogische Psychologie, 25*(1), 5–18.

Dalbert, C. (2013). *Gerechtigkeit in der Schule [Justice in School]*. Wiesbaden: Springer.

Dalbert, C., & Stoeber, J. (2005). The belief in a just world and distress at school. *Social Psychology of Education, 8*(2), 123–135.

Deutsch, M. (1985). *Distributive justice: A social-psychological perspective*. New Haven: Yale University Press.

Donat, M., Dalbert, C., & Kamble, S. V. (2014). Adolescents' cheating and delinquent behavior from a justice-psychological perspective: The role of teacher justice. *European Journal of Psychology of Education, 29*(4), 635–651. https://doi.org/10.1007/s10212-014-0218-5.

Donat, M., Herrmann, M., & Umlauft, S. (2013). Gerechtigkeitserleben und Sozialverhalten von SchülerInnen [Justice experience and social behavior of pupils]. In C. Dalbert (Ed.), *Gerechtigkeit in der Schule* (pp. 73–92). Wiesbaden: Springer VS.

Ehrhardt, N., Pretsch, J., Herrmann, I., & Schmitt, M. (2016). Observing justice in the primary school classroom. *Zeitschrift Für Erziehungswissenschaft, 19*(1), 157–190. https://doi.org/10.1007/s11618-015-0664-0.

Folger, R. G., & Cropanzano, R. (1998). *Organizational justice and human resource management*. Thousand Oaks, CA: SAGE Publications.

Furnham, A., & Siegel, E. M. (2012). Reactions to organizational injustice: Counter work be-haviors and the insider threat. In E. Kals & J. Maes (Eds.), *Justice and conflicts* (pp. 199–217). Berlin: Springer.

Gollwitzer, M., & van Prooijen, J.-W. (2016). Psychology of Justice. In C. Sabbagh & M. Schmitt (Eds.), *Handbook of social justice theory and research* (pp. 61–82). New York, NY: Springer. https://doi.org/10.1007/978-1-4939-3216-0_4.

Goodman, R. (1997). The strengths and difficulties questionnaire: A research note. *Journal of Child Psychology and Psychiatry, 38,* 581–586.

Greenberg, G. (1990). Employee theft as a reaction to underpayment inequity: The hidden cost of pay cuts. *Journal of Applied Psychology, 75,* 561–568.

Greenberg, J., & Colquitt, J. (Eds.). (2005). *Handbook of organizational justice*. Mahwah, N.J: Lawrence Erlbaum Associates.

Hagenauer, G. (2011). *Lernfreude in der Schule*. Waxmann: Münster u.a.

Hascher, T., & Edlinger, H. (2009). Positive Emotionen und Wohlbefinden in der Schule - Ein Überblick über Forschungszugänge und Erkenntnisse [Positive emotions and well-being in school—A review of research approaches and findings]. *Psychologie in Erziehung und Unterricht, 56*(2), 105–122.

Helmke, A. (1993). Die Entwicklung der Lernfreude vom Kindergarten bis zur 5. Klassenstufe [The development of joy of learning in children from kindergarten to grade five]. *Zeitschrift für Pädagogische Psychologie, 7,* 77–86.

Higgins, E. T., & King, G. A. (1981). Accessibility of social constructs: Information processing consequences of individual and contextual variability. In N. Cantor & J. F. Kihlstrom (Eds.), *Personality, cognition, and social interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Israelashvili, M. (1997). Situational determinants of school students' feelings of injustice. *Elementary School Guidance & Counseling, 31*(31), 283–292.

Jasso, G., Törnblom, K. Y., & Sabbagh, C. (2016). Distributive justice. In C. Sabbagh & M. Schmitt (Eds.), *Handbook of social justice theory and research* (pp. 201–218). New York, NY: Springer. https://doi.org/10.1007/978-1-4939-3216-0_11.

Klasen, H., Woerner, W., Wolke, D., Meyer, R., Overmeyer, S., Kaschnitz, W., et al. (2000). Comparing the German versions of the Strengths and Difficulties Questionnaire (SDQ-Deu) and the Child Behavior Checklist. *European Child and Adolescent Psychiatry, 9*(4), 271–276.

Kline, R. B. (2015). *Principles and practice of structural equation modeling (4th Revised edition)*. New York: Taylor & Francis Ltd.

Lerner, M. J. (1977). The justice motive: Some hypotheses as to its origins and forms. *Journal of Personality, 45*(1), 1–52.

Leventhal, G. S. (1980). What should be done with equity theory? In K. J. Gergen, M. S. Greenberg, & R. H. Willis (Eds.), *Social exchange* (pp. 27–55). New York: Springer.

Lind, E. A., & Tyler, T. R. (1988). *The social psychology of procedural justice*. Boston, MA: Springer. https://doi.org/10.1007/978-1-4899-2115-4.

Lotz, M., Gabriel, K., & Lipowsky, F. (2013). Niedrig und hoch inferente Verfahren der Unterrichtsbeobachtung. Analysen zu deren gegenseitiger Validierung [Low and high inference rating instruments for class observation]. *Zeitschrift für Pädagogik, 59*(3), 357–380.

Maltese, S., Baumert, A., Knab, N., & Schmitt, M. J. (2013). Learning to interpret one's own outcome as unjustified amplifies altruistic compensation: A training study. *Personality and Social Psychology, 4,* 951. https://doi.org/10.3389/fpsyg.2013.00951.

Molinari, L., Speltini, G., & Passini, S. (2013). Do perceptions of being treated fairly increase students' outcomes? Teacher–student interactions and classroom justice in Italian adolescents. *Educational Research and Evaluation, 19*(1), 58–76. https://doi.org/10.1080/13803611.2012.748254.

Parsons, T. (1959). The school class as a social system: Some of its functions in American society. *Harvard Educational Review, 29*(4), 297–318.

Paulsel, M. L., & Chory-Assad, R. M. (2005). Perceptions of instructor interactional justice as a predictor of student resistance. *Communication Research Reports, 22*(4), 283–291.

Pekrun, R. (1993). Facets of students' academic motivation: A longitudinal expectancy-value approach. In M. Maehr & P. Pintrich (Eds.), *Advances in Motivation and Achievement* (Vol. 8, pp. 139–189). Greenwich, CT: JAI Press.

Peter, F., & Dalbert, C. (2010). Do my teachers treat me justly? Implications of students' justice experience for class climate experience. *Contemporary Educational Psychology, 35,* 297–305.

Piaget, J. (1997). *The moral judgment of the child*. New York: Free Press Paperbacks.

Pornprasertmanit, S., Miller, P., Schoemann, A., & Rosseel, Y. (2014). semTools: Useful tools for structural equation modeling. R package version 0.4-6. http://CRAN.R-project.org/package=semTools.

Pretsch, J., Ehrhardt, N., Engl, L., Risch, B., Roth, J., Schumacher, S., et al. (2015). Injustice in school and students' emotions, well-being, and behavior: A longitudinal study. *Social Justice Research, 29*(1), 119–138. https://doi.org/10.1007/s11211-015-0234-x.

R Core Team. (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. http://www.R-project.org/.

Rauer, W., & Schuck, K. D. (2003). Fragebogen zur Erfassung emotionaler und sozialer Schulerfahrungen von Grundschulkindern dritter und vierter Klassen (FEESS 3–4). Göttingen: Hogrefe; *Praxis Der Kinderpsychologie Und Kinderpsychiatrie, 53*(2), 139–141.

Rauer, W., & Schuck, K. D. (2004). *FEESS 1-2 Fragebogen zur Erfassung emotionaler und sozialer Schulerfahrungen von Grundschulkindern erster und zweiter Klassen [Questionnaire to assess emotional and social experiences in the first two years of primary school]*. Göttingen: Beltz.

Reinke, W. M., Herman, K. C., & Newcomer, L. (2016). The Brief Student-Teacher Classroom Interaction Observation: Using dynamic indicators of behaviors in the classroom to predict outcomes and inform practice. *Assessment for Effective Intervention*. https://doi.org/10.1177/1534508416641605.

Resh, N., & Sabbagh, C. (2014). Justice, belonging and trust among Israeli middle school students. *British Educational Research Journal, 40*(6), 1036–1056.

Resh, N., & Sabbagh, C. (2016). Justice and education. In C. Sabbagh & M. Schmitt (Eds.), *Handbook of social justice theory and research* (pp. 349–367). New York, NY: Springer. https://doi.org/10.1007/978-1-4939-3216-0_19.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36. http://www.jstatsoft.org/v48/i02/.

Rudasill, K. M., Rimm-Kaufman, R. E., Justice, L. M., & Pence, K. (2006). Temperament and language skills as predictors of teacher–child relationship quality in preschool. *Early Education & Development, 17*(2), 271–291. https://doi.org/10.1207/s15566935eed1702_4.

Sabbagh, C., & Resh, N. (2014). Citizenship orientations in a divided society: A comparison of three groups of Israeli junior-high students-secular Jews, religious Jews, and Israeli Arabs. *Education, Citizenship and Social Justice, 9*(1), 34–54.

Sabbagh, C., & Schmitt, M. (Eds.). (2016). *Handbook of social justice theory and research*. New York, NY: Springer. https://doi.org/10.1007/978-1-4939-3216-0.

Schmitt, M., Baumert, A., Gollwitzer, M., & Maes, J. (2010). The justice sensitivity inventory: Factorial validity, location in the personality facet space, demographic pattern, and normative data. *Social Justice Research, 23*(2–3), 211–238. https://doi.org/10.1007/s11211-010-0115-2.

Schmitt, M., & Dörfel, M. (1999). Procedural injustice at work, justice sensitivity, job satisfaction and psychosomatic well-being. *European Journal of Social Psychology, 29*(4), 443–453. https://doi.org/10.1002/(SICI)1099-0992(199906)29:4<443:AID-EJSP935>3.0.CO;2-C.

Schmitt, M., Gollwitzer, M., Maes, J., & Arbach, D. (2005). Justice sensitivity. *European Journal of Psychological Assessment, 21*(3), 202–211. https://doi.org/10.1027/1015-5759.21.3.202.

Schmitt, M., Rebele, J., Bennecke, J., & Förster, N. (2008). Ungerechtigkeitssensibilität, Kündigungs-gerechtigkeit und Verantwortlichkeitszuschreibungen als Korrelate von Einstellungen und Verhalten Gekündigter gegenüber ihrem früheren Arbeitgeber (Post Citizenship Behavior). [Justice sensitivity, justice of termination and attribution of responsibility as correlates of attitudes and behavior of terminated employees towards their former employer]. *Wirtschaftspsychologie, 10,* 101–110.

Shaw, D. S., Gilliom, M., Ingoldsby, E. M., & Nagin, D. S. (2003). Trajectories leading to school-age conduct problems. *Developmental Psychology, 39*(2), 189–200. https://doi.org/10.1037/0012-1649.39.2.189.

Skalická, V., Belsky, J., Stenseng, F., & Wichstrøm, L. (2015). Reciprocal relations between student-teacher relationship and children's behavioral problems: Moderation by child care group size. *Child Development, 86*(5), 1557–1570. https://doi.org/10.1111/cdev.12400.

Susteck, H. (1996). Rituale in der Schule [Rituals in school]. *Pädagogische Welt, 50*(1), 34–38.

Tacke, M. (2006). Wohlbefinden in der Schule - Ein Beitrag zur gesunden Schule. [Well-being in school—A contribution to a healthy school]. In *Vom Nürnberger Trichter zum Laptop? Schule zwischen kognitivem und sozial-emotionalem Lernen. Kongressbericht der 16. Bundeskonferenz 2004 in Nürnberg* (pp. 694–702). Bonn: Deutscher Psychologen Verlag.

Tata, J. (1999). Grade distributions, grading procedures, and students' evaluations of instructors: A justice perspective. *The Journal of Psychology, 133*(3), 263–271.

Terwee, C. B., Bot, S. D. M., de Boer, M. R., van der Windt, D. A. W. M., Knol, D. L., Dekker, J., et al. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology, 60*(1), 34–42. https://doi.org/10.1016/j.jclinepi.2006.03.012.

Thorkildsen, T. A. (1989). Justice in the classroom: The student's view. *Child Development, 60*(2), 323. https://doi.org/10.2307/1130079.

Van den Bos, K., Maas, M., Waldring, I. E., & Semin, G. R. (2003). Toward understanding the psychology of reactions to perceived fairness: The role of affect intensity. Retrieved August 24, 2016, from http://dspace.library.uu.nl/handle/1874/7851.

Vermunt, R., & Steensma, H. (2016). Procedural justice. In C. Sabbagh & M. Schmitt (Eds.), *Handbook of social justice theory and research* (pp. 219–236). New York, NY: Springer. https://doi.org/10.1007/978-1-4939-3216-0_12.

Wenzel, M., & Okimoto, T. G. (2016). Retributive Justice. In C. Sabbagh & M. Schmitt (Eds.), *Handbook of social justice theory and research* (pp. 237–256). New York, NY: Springer. https://doi.org/10.1007/978-1-4939-3216-0_13.

Whisenant, W., & Jordan, J. S. (2008). Fairness and enjoyment in school sponsored youth' sports. *International Review for the Sociology of Sport, 43*(1), 91–100. https://doi.org/10.1177/1012690208093470.

Wustmann Seiler, C. (2012). Erfassung des schulischen Wohlbefindens am Schulanfang: Empirische Überprüfung eines mehrdimensionalen Konstrukts [Assessing well-being in school at the beginning of primary school]. *Zeitschrift für Grundschulforschung, 5*(2), 136–150.

Natalie Ehrhardt-Madapathi is a postdoctoral researcher at the Department of Differential Psychology and Psychological Assessment of the University of Koblenz-Landau. Her main research interests are justice sensitivity, classroom justice, inclusive school systems and the role of teacher training in teacher justice.

Johanna Pretsch is a postdoctoral researcher at the Department of Differential Psychology and Psychological Assessment of the University of Koblenz-Landau. Her main research interests are personality assessment, individual differences in education, classroom justice, and the development of democratic attitudes in students.

Manfred Schmitt is Professor of Differential Psychology and Psychological Assessment at the University of Koblenz-Landau. He is interested in social justice, emotion, justice sensitivity and prosocial behavior.