

Observing justice in the primary school classroom

Natalie Ehrhardt · Johanna Pretsch · Ivana Herrmann · Manfred Schmitt

Published online: 18 December 2015
© Springer Fachmedien Wiesbaden 2015

Abstract Justice-related situations are a part of students' everyday life. In order to test the antecedents, correlates, and consequences of (in)justice in school, valid measures of justice are needed. To our knowledge, this is the first study to develop an observer low inference rating instrument that can be applied to measure justice in the primary classroom. In two pre-studies, justice-relevant situations in the classroom were extracted and observable indicators for these situations were developed. In the main study, this instrument was used to observe 208 primary school students with regard to their experiences of justice or injustice. In addition to this, other measures of justice were developed to examine the convergence between observer low inference ratings of classroom justice and high inference rating instruments for teachers, students, and external observers.

Factor analyses and correlations between the different indices of the observer low inference rating and the high inference rating items suggested that incidents of justice and injustice in the classroom do not tend to co-occur frequently. Teachers do not appear to have a general tendency to treat a child more or less justly across a large number of situations.

N. Ehrhardt (✉) · I. Herrmann
Graduate School "Teaching and Learning Processes", University of Koblenz-Landau,
Thomas-Nast-Straße 44,
76829 Landau, Germany
e-mail: ehrhardt@uni-landau.de

I. Herrmann
e-mail: herrmann@uni-landau.de

Dr. J. Pretsch · Prof. Dr. M. Schmitt
Department of Psychology, University of Koblenz-Landau,
Fortstr. 7,
76829 Landau, Germany
e-mail: pretsch@uni-landau.de

Prof. Dr. M. Schmitt
e-mail: schmittm@uni-landau.de

The findings suggest that a comprehensive assessment of classroom justice requires a multi-method approach where the justice ratings of students, teachers and external observers are all taken into consideration.

Keywords Justice · Primary school · Observation

Beobachtung von Unterrichtsgerechtigkeit in der Grundschule

Zusammenfassung Gerechtigkeitsrelevante Situationen gehören zum Schulalltag. Um die Auswirkungen von (un)gerechter Behandlung auf Schüler/innen untersuchen zu können, sind valide Instrumente notwendig. In dieser Studie wurde ein Instrument zur systematischen Beobachtung der Unterrichtsgerechtigkeit in der Grundschule entwickelt. In zwei Vorstudien wurden gerechtigkeitsrelevante Situationen extrahiert und beobachtbare Indikatoren für diese Situationen entwickelt. In der Hauptstudie fand das niedrig inferente Ratinginstrument an 208 Grundschüler/inne/n Anwendung. Des Weiteren wurden hoch inferente Ratinginstrumente zur Unterrichtsgerechtigkeit aus Sicht der Lehrperson, der Schüler/innen und der externen Beobachter entwickelt und angewendet, um die Konvergenz dieser unterschiedlichen Instrumente der Unterrichtsgerechtigkeit zu untersuchen.

Faktoranalysen der niedrig und der hoch inferenten Ratinginstrumente deuten darauf hin, dass verschiedene gerechte oder ungerechte Ereignisse im Unterricht nicht häufig gleichzeitig auftreten. Lehrpersonen haben anscheinend nicht die Tendenz, ein Kind pauschal über verschiedene gerechtigkeitsrelevante Situationen hinweg mehr oder weniger gerecht zu behandeln.

Die Ergebnisse sprechen dafür, dass Schulgerechtigkeit am besten durch einen Multi-Trait-Multi-Method Ansatz gemessen werden kann, in welchem die Ratings von Lehrer/inne/n, Schüler/inne/n und externen Beobachtern Berücksichtigung finden.

Schlüsselwörter Beobachtung · Gerechtigkeit · Grundschule

1 Introduction

1.1 The importance of justice

Many important topics in people's lives are related to justice in one way or another, for example, how to treat people of a different religion adequately, how to allocate money and wages fairly, and how to construct a government that ensures social justice. Schwan (2008) stated that justice is the leading principle in human coexistence. Because humans do not live their lives separate from each other, they cannot avoid social interactions. Hence, every society has to consider how to make interactions fair and how to distribute goods fairly.

A primary school classroom as a small society in itself has to face the same problems. For example, teachers must consider how to allocate time and attention to stu-

dents justly, how to interact with students in a fair and respectful way, how to cope with disruptions in a fair way, or how to develop rules that ensure fair learning conditions for everyone. When a situation is perceived as unjust, people react with negative emotions because they have a need for justice.

According to justice theories such as equity theory (Adams 1965) and relative deprivation theory (Crosby 1984), people evaluate the justice of a situation on the basis of their outcomes in relation to the outcomes of others as well as on everyone's respective inputs. People want to get what they deserve and deserve what they get (Montada and Lerner 1998). However, what someone deserves is a question of the principle of justice that is applied to the specific issue. What justice principles are there? And what do these principles define as just?

1.2 Definitions of justice (forms and principles)

Early approaches to the question of what is just stem from Aristotle, who defined justice as the equal treatment of equals (Aristotle 1998). A recent definition of justice was proposed by Lerner (Lerner 1977, 1980) who argued that justice is served if everyone gets what he or she deserves. Both definitions can be applied to a virtually unlimited number of situations that individuals and groups encounter in their lives.

Given the eminent role justice plays in social interactions and the many situations in which people request justice, it comes as no surprise that justice is a core issue in various scientific disciplines. Philosophy, jurisprudence, and political science approach justice in a normative way. They reflect on and discuss what a just society should look like (Meyer and Sanklecha *in press*; Rawls 1971). Social sciences such as psychology and sociology as empirical sciences do not try to define what is truly just or truly unjust. Rather, they aim to determine what people regard as just, whether individuals differ in their justice perceptions, how people react to injustice, how justice judgments depend on the social context, and to what extent justice judgments converge across the actors and observers who are involved in a justice case (Gollwitzer and Van Prooijen *in press*; Liebig and Sauer *in press*). The present research addresses some of these empirical questions in an educational context: the classroom.

1.2.1 *Forms of justice*

Four forms of justice have been differentiated in the recent justice literature (Sabbagh and Schmitt *in press*): distributive justice, retributive justice, procedural justice, and interactional justice.

The question of *distributive justice* arises when there is more than one party and a limited resource or desired good. Distributive justice considers whether a distribution of goods is fair. The goods that are distributed do not have to be real objects; they can just as well be time, attention, or praise (Jasso et al. *in press*).

Individuals, groups, and societies also distribute punishment. In this case, principles of *retributive justice* are relevant. Retributive justice means that a wrong-doing is followed by a punishment that fits the offense. Principles of retributive justice typically consider either the amount of harm the transgressor has caused or the number of

advantages the transgressor has gained from the transgression (Wenzel and Okimoto [in press](#)).

Procedural justice is related to the processes that lead to the outcome such as a distribution of goods or punishment (Leventhal 1980; Lind and Tyler 1988). In school, for instance, procedural justice can mean that a student understands a teacher's grading process and considers the process to be fair even when the student receives a bad grade (Tata 1999).

Interactional justice refers to the fairness and quality of interactions (Bies and Moag 1986). Interactions are fair when people are sensitive, kind, and respectful to each other.

1.2.2 Justice principles

The different forms of justice that we introduced define which kinds of actions and interactions are justice-relevant. In order to judge the decision or outcome of a specific situation, these forms have to be combined with principles. The most important principles of distributive justice are equity, equality, and need (Deutsch 1985; Jasso et al. [in press](#)). According to Deutsch (1985), people do not decide randomly which principle to apply in which situation, but there is a fit between the situation and the justice principle of choice.

The *equality* principle requires everyone to be treated equally and to receive the same outcome. This principle is considered most appropriate in small intimate groups such as families (Deutsch 1985).

In contrast to equality, the justice principle of *equity* requires a differentiation (Jasso et al. [in press](#)). Equity justice is served if not everyone receives the same treatment, but everybody receives what he or she deserves with respect to the input that he or she provided. The equity principle is considered most appropriate in economic contexts and contexts that are defined by competition such as sports (Deutsch 1985). For example, a person who produces a better product should therefore receive a higher wage than a less successful coworker.

The *need* principle also requires differentiation (Jasso et al. [in press](#)). Here, everyone should get what he or she needs in order to have a decent life. The need principle is often favored in caring-oriented group contexts (Deutsch 1985). One example of such a context is a primary school class.

1.3 Justice in school

1.3.1 The importance of justice in school

School is an important institution of socialization. It is the first institution that children have to attend and where they have to follow the rules of this organization (Correia et al. 2009). Experiencing justice in this first institution might provide a foundation for organizational trust in children (Resh and Sabbagh [in press](#)). A child's first years in school may be especially crucial for this.

Schools are expected to help children integrate into society and to foster their social development (Susteck 1996). To achieve this goal, the classroom has to pro-

vide a caring-oriented social environment that considers children's needs. In justice-related situations, this can be achieved by applying the need principle (Deutsch 1985; Berti et al. 2010). The teacher as the most powerful person in a classroom has a great responsibility to make sure that every child is granted fair treatment and that every child's needs are met (Sabbagh and Resh 2014). The teacher has to decide how to allocate rewards, attention, and punishment (Connell 1993). The teacher also has to evaluate students' performance and their social behavior (Resh and Sabbagh 2014). Besides the teacher, peers provide another important source of justice experiences. According to Petillon (1993), peers are the main source of perceived unjust treatment in first graders. The (un)fairness of interactions with peers has a substantial influence on a child's development. Moreover, discussions of fairness issues can also contribute to the social development of children. Damon and Killen (1982) found that peer discussions about a social justice topic improved children's moral development more than a similar discussion with an adult did.

Even though there is consensus among students and teachers on the importance of justice in the classroom (Kanders 2000), half of the students from several countries have claimed that they have experienced injustice in school (Israelashvili 1997; Dalbert 2011).

1.3.2 *Effects of injustice in school*

Given the importance of justice, research on the effects of (un)fairness in the classroom is surprisingly scarce. One of the earliest studies revealed that students who perceived their teacher as fair also liked the teacher (Tata 1999). Students who rated their teacher as fair were also more motivated to learn and less inclined to exhibit aggressive behavior during lessons (Chory-Assad 2002; Paulsel and Chory-Assad 2005). Students who perceived their teacher as interacting fairly reacted positively to requests and were more helpful (Wubbels and Brekelmans 2005). These and additional findings (Dalbert 2011; Pretsch et al. *in press*) suggest that justice in school promotes students' pro-social behavior and their subjective well-being, whereas injustice has detrimental effects on students' social behavior and their emotional well-being.

1.4 Measuring justice in school

In order to test the antecedents, correlates, and consequences of (in)justice in school, valid measures of justice are needed. This task is challenging because justice perceptions and judgments depend on which role the perceiver plays in a justice-relevant situation (Mikula et al. 1990). Up to four typical roles can be involved in a justice incident. A person can be the recipient of a distribution and feel as a victim if treated unfairly, a person can be the allocator of a good or a punishment and feel as a perpetrator if acting unfairly, a person can passively benefit from an unfair distribution and feel as a beneficiary. A person can also observe a justice relevant incident without being actively involved and affected by a decision. Such observers can serve as neutral judges if they do not identify with one of the involved parties (Baumert and Schmitt *in press*). Although judges have subjective views on a justice case, their judgments are less biased

by conflicting motives such as self-enhancement and social desirability concerns. The justice judgments of neutral observers may come closest to “objective” justice.

A comprehensive measure of justice in the classroom has to consider as many perspectives as possible. Previous studies have measured justice from only teachers’ or students’ subjective points of view (e.g., Dalbert and Stoeber 2005; Correia and Dalbert 2007). To the best of our knowledge, neutral observers have not been previously employed in the measurement of classroom justice. Our current study was designed to fill this gap. Our primary goal was to develop a comprehensive measure of first-grade classroom justice and to test the degree of convergence between its components. Special attention was given to the development of an observational measure to be used by neutral observers because such a measure is still missing and comes closest to what might be considered “objective” justice.

1.4.1 What is classroom justice?

Fair treatment is not general but is bound to a context (Deutsch 1985). In order for a child to receive fair treatment in the classroom, the class rules have to be fair, and the people involved in the interactions have to act fairly. An instrument for measuring classroom justice needs to provide a measure of fair treatment on the level of an individual child rather than on the class level. This is true because it is possible and rather likely that whereas one particular child’s needs may be met and he or she may be treated fairly, another child in the same classroom may be neglected.

As mentioned earlier, we consider need to be the justice principle of choice in first-grade classrooms. The equality principle does not seem to be appropriate at this early stage in the educational process because children enter first grade with very different skills and abilities. Therefore, they do not have the same opportunities to learn or grow if they receive equal treatment in school. For the same reason, the equity principle is also not appropriate as a principle of distributive justice in first-grade classroom contexts.

Rather than treating children strictly equally or primarily based on their achievements, a teacher has to be sensitive to the many needs that children have when entering school. A child in primary school has a need for secure emotional attachment and positive regard by his or her teacher and classmates. The child also has a need to learn basic academic skills such as reading and writing. In order to improve their academic skills, children need time to work on tasks, opportunities to speak (Corden 2000), appropriate feedback (Burnett and Mandel 2010), emotional support (Osterman 2000), encouragement, and praise.

Applying the principle of need requires differentiation. We would consider it fair, for instance, if a child who did not understand the task was given extra attention and support from the teacher in order to reach his or her learning goal, while a child who already understood the task was left to solve the task without any help from the teacher. The same rationale applies to praise. A child who has worked very hard to accomplish a task might be in need of praise, whereas another child for whom the task was not challenging would not profit from praise. Praise for solving an easy task might even have a detrimental effect on the child’s motivation and self-efficacy (Henderlong and Lepper 2002). In their review Stroet et al. (2013) concluded that

there was a consistent relationship between need supportive teaching and early adolescents' motivation and engagement for school.

These examples show that applying the need principle as the most appropriate justice principle in primary school classrooms makes measuring classroom justice more complex than applying the equality principle. In order to observe and rate a treatment as fair according to the need principle, it is not sufficient to consider only outcomes of allocations such as the allocation of attention, time, support, and praise. Rather, these outcomes must be compared with the child's needs. Therefore, the child's needs have to be assessed too, and this in turn means that need indicators must be defined. For example, a child can indicate his or her need by asking for help or raising his or her hand to indicate the need to speak. Even if a child is not asking explicitly, a bad task performance could also be a need indicator for greater teacher support. Moreover, a child's reaction to teacher support can also indicate the child's need for support. If the child reacts positively to help, then he or she needed help. If the child reacts negatively to help, then he or she either had no need for help or else was not given the right help.

1.4.2 Methods for measuring classroom justice

Such indicators of the child's needs can be used by observers. Teachers, as actors and potential perpetrators, also have to rely on students' need indicators. In addition, they have to self-observe their own behavior in order to judge how well they met a child's needs. Because teachers are actively involved in the teaching process and because they interact with many students, they cannot rely on strict rules for recognizing need indicators and their behaviors. Rather, they have to recollect relevant information and integrate it into a summary of high inference rating after a lesson. Thus, although teachers and observers alike observe need indicators expressed by students and teachers' behaviors, their tasks in measuring justice differ. Teachers can only give high inference justice ratings, but external observers can observe justice indicators as low inference ratings and additionally provide high inference ratings on this basis.

Students as the targets of teachers' behaviors are also actively involved in the learning process during a class. Therefore, they cannot be requested to provide protocols of any sort. However, students are clearly the experts of their needs, and they are the ones who will have feelings about whether the treatment they receive from the teacher and their peers is fair or unfair. Thus, in principle, they are a valuable source of information when it comes to measuring classroom justice.

Student high inference ratings Are primary school students able to make justice judgments? Secondary school students are able to assess complex constructs such as teachers' instructional quality (Gruehn 2000; Ditton 2002; Hattie 2009). But primary school students still struggle to provide valid answers to interview questions with an open answer format due to their limited vocabulary, their limited conceptual understanding, and because they have trouble aggregating and abstracting information from single incidents (Piaget 1997; Biemer and Lyberg 2003; Goswami 2011).

Because classroom justice is a complex phenomenon to judge, students as judges have to integrate different aspects of a situation. Moreover, they have to consider the different perspectives of the various actors. Research in moral development and

social cognition (e.g., Harris 1989) has shown that this is very difficult for children younger than 8 years. Younger children are not able to seize all relevant information and often focus on a single aspect instead. Piaget (1997) argued that children up to the age of 7 or 8 years believe in immanent justice. From this point of view, it can be expected that, as in any other area of their lives, younger children will judge the classroom to be fair because they do not believe that justice is something that has to be actively produced.

These theoretical considerations and findings suggest that justice judgments provided by children at the beginning of primary school may not converge with judgments of other perceivers. We tested this hypothesis in the present research by comparing the justice judgments of students with those of teachers and observers.

Teacher high inference ratings Teachers are expected to know their students very well, and this includes their students' needs. Accordingly, Honkanen et al. (2014) showed that teachers were able to assess their students' mental health rather precisely and were even able to predict future mental health problems. In addition to their knowledge of the students, teachers also have a lot of experience with classroom justice (Kanders 2000). Although teachers are experts on their students, their high inference rating of classroom justice may be impaired. As Au et al. (2007, p. 10) stated, "Teachers are often simultaneously perpetrators and victims, with little control over planning time, class size, or broader school policies." This ambiguous role of observing injustice, suffering from injustice as a victim, and committing injustice could decrease the accuracy of teachers' assessments of classroom justice. In addition, it is a very extensive task for the teacher to judge justice in the classroom for every single child. For these reasons, teacher high inference ratings of classroom justice may differ from high inference rating provided by neutral observers. We tested this possibility in the present study by comparing teacher high inference ratings with high inference ratings made by neutral observers.

Observer low inference and high inference ratings As stated earlier, uninvolved neutral observers are probably able to provide the most "objective" evaluations of classroom justice. A promising method might be systematic behavioral observation in a naturalistic classroom setting during a lecture when all students and the teacher are present. Naturalistic observations should provide an ecologically valid measure of justice in the classroom. Moreover, they avoid the ethical pitfalls of a lab experiment because instances of injustice appear in the classroom without the researchers' interference (Dalbert 2011). By observing the child's needs and how these needs are met, it is possible to consider the individuality of the fairness that each child receives. Even without the expert knowledge of the teacher and the insights of the students themselves, a neutral observer can rate the appropriateness of interactions and thus the degree of fair treatment in the classroom.

1.5 Which forms of justice are observable in the classroom?

Some justice-relevant events and behaviors in the classroom are discrete, obvious, and can be observed directly such as the feedback a child receives from the teacher.

Other justice-relevant events and behaviors, such as adaptive teaching, are complex processes that cannot easily be decomposed into discrete observable elements. Rather, the elements of such complex processes have to be integrated mentally by observers and transformed into a justice rating.

Retributive justice was described earlier as the question of whether the punishment fits the offense. Teachers sometimes punish their students, and although the punishment itself is observable, determining the appropriateness of a punishment is a complex process. Single events that could be quantified and counted by observers would not provide adequate measurement; rather, a combination of the situation, the punishment, and the student's reaction to the punishment have to be mentally integrated by the observers and aggregated into a justice rating of retributive justice.

Procedural justice is difficult to observe and especially difficult to quantify. An observer cannot count how often procedural justice occurs in school. In many situations, an external observer cannot judge procedural justice because he or she has no knowledge of the underlying processes, which are not always observable. If, for example, a teacher asks one student to perform a task that all of the students like a lot, then the observer does not know whether the teacher has allocated the popular tasks in a fair way or not. Sometimes the teacher may explicitly explain to the children the process of choosing who gets to perform which task. In this case, an observer could judge the procedural justice of the teacher on the basis of the importance he or she gives to the observed incidents. But the teacher might often use a fair process for allocating tasks without explicitly explaining the process every time. Still, there are indicators that can be observed or rated. One example is the reaction of the teacher to a child who yells out an answer even though it was not his or her turn. The teacher has three options. He or she could scold the child, could accept the child's remark, or could ignore the child's remark. The fairest process would be to ignore the child's remark. If the teacher scolds the child, then this itself disrupts the lecture and gives the child attention that he or she does not deserve at this moment. If the teacher accepts the child's remark, this would be unfair to the other children. It would also be unfair to the child who is the perpetrator because the child needs to learn that he or she cannot disrupt the lecture, and for that, he or she should not be given attention for his or her disruption. The way in which the teacher reacts to disturbances is therefore an observable aspect of procedural justice. Other aspects of procedural justice might be evaluated better by computing a mental summary of events so that several behaviors can be considered together. This summary can result in high inference ratings of procedural justice. There are different aspects of procedural justice that can be rated, for example: Was the student given enough time to think? Did the learning procedure enable individual learning?

Distributive justice can be observed rather easily in class. In particular, the allocation of time and attention dedicated to each student by the teacher can be observed and even quantified. Other aspects of distributive justice might be observable but might occur less frequently and are therefore not well-suited for direct observation (e.g., the distribution of desirable objects such as birthday cakes). The fair distribution of free choices and restrictions to freedom are complex processes that cannot be easily decomposed into discrete observable elements. Rather, the elements of such complex processes have to be integrated mentally by observers and transformed into a justice rating.

Interactional justice can also be observed rather well. Even if observers know little about the students and the teacher, they can observe the interactions and judge the adequacy of such interactions. They can use verbal as well as nonverbal cues (e.g., tone of voice). Interactional justice can be observed in the classroom by both direct observations of single incidents and a mental summary of events and multiple behaviors. Some of the single incidents that can be observed are the number of times a teacher actively listens to a child, the number of times a child is disturbed or hit by another child, the number of times a child is denied help after asking for it. The following aspects of interactional justice are better suited to being rated than observed directly: The tone of the interaction was respectful, praise and criticism were constructive, the tone between the peers and the child was respectful, and so forth.

In conclusion, it is reasonable to assume that interactional justice and distributive justice can be observed through single incidents in the classroom, whereas it is more difficult to observe retributive and procedural justice. For these justice forms, high inference ratings might be more appropriate. But in order to measure justice in the classroom, it might not be necessary to measure all of the different forms. Colquitt et al. (2001) showed in their meta-analytic review of organizational justice that the different forms of justice are distinct but still share a substantial amount of variance. However, it might also be possible that the various ways in which students can be treated fairly or unfairly in class do not co-occur. Rather, they may be distinct and occur independently from each other. For example, a teacher might treat a specific child in class with respect (procedural fairness) but fail to meet the child's need for attention, appropriate feedback, and praise (distributive justice). Because nothing is known so far about the degree of overlap between forms of justice in the classroom, our study also addressed this issue.

1.6 Aim of the present study

The primary goal of this study was to develop an observer low inference rating instrument to measure justice in its different forms in first-grade classrooms. Since to our knowledge, classroom justice has not yet been measured in such a comprehensive way, a second goal was to uncover the factorial structure of the forms and indicators of classroom justice. Specifically, we wanted to explore the degree to which different kinds of treating students fairly co-occur. Third, we wanted to explore the extent of convergence between observer low inference and high inference ratings of classroom justice in order to find out whether high inference ratings can substitute the more laborious observer low inference ratings. For this reason, we also developed high inference justice-rating instruments for observers, teachers, and students. Fourth, we wanted to identify the amount of convergence among the justice ratings as provided by the different persons who are involved or observe justice relevant interactions between teachers and students and among students. High convergence would mean that students, teachers, and neutral observers have consensus regarding the fair treatment of students in the classroom. Low convergence would mean that justice is strongly in the eye of the beholder and reflects subjective perceptions and evaluations of events that cannot be generalized across the perspectives on and the roles played in justice relevant situations.

2 Method

2.1 Development of the observer low inference and the high inference rating instruments

To measure justice along the different justice forms in a caring-oriented primary school classroom setting, it is necessary to assess the needs of the individual child and also how well these needs are met. This cannot be achieved without considering the context because a child's needs can vary from one situation to another. Therefore, it is necessary to identify situations in which classroom justice can be either observed or rated or both. In order to identify such situations, we conducted two pre-studies, the first with primary school teachers, and the second with first-grade students.

2.1.1 Pre-study 1

In Pre-study 1, 11 primary school teachers and one secondary school teacher (11 women, mean age: 38 years, mean teaching experience: 8.3 years) answered open-ended questions via an online survey tool. Teachers were asked four questions: (1) What are typical justice-related situations in the classroom, i.e., situations in which students typically feel that they have been treated either fairly or unfairly? (2) In what situations do you struggle as a teacher with the challenge of treating students fairly? (3) The third question was preceded by definitions of equality, equity, and need as the main principles of distributive justice. Subsequently, teachers were asked to give examples of classroom situations in which these principles are relevant. (4) The fourth question was preceded by definitions of the main justice forms we introduced earlier. Subsequently, teachers were asked to provide examples of classroom situations in which these justice forms are relevant.

With a content analysis (Mayring 2010), the total of 17 distinct kinds of situations described by the teachers were classified according to the justice forms and principles of distributive justice. The content analysis revealed that all situations described by the teachers could be differentiated according to distributive, retributive, procedural, and interactional justice. The content analysis also revealed that the principles of distributive justice discriminated less well between justice-related situations than the justice forms did. In other words, typical justice situations in the classroom can be characterized better by the justice form they represent (distributive, procedural, interactional, retributive) than by the distributive justice principle that is considered appropriate in the situation (equity, equality, need). Whereas the justice situations were rather evenly distributed across the justice forms, they were unevenly distributed across the justice principles because most teachers gave priority to the need principle over the other principles. This is in line with work by Deutsch (1985) and our previously introduced assumption that primary school is perceived more as a caring-oriented social context than as an achievement-oriented context.

Table 1 describes typical situations (Column 2) that can be mapped onto the four justice forms (Column 1). The table also gives the total number of situations of each type that were generated by the teachers (Column 3) and the number of teachers who gave at least one example of the situation in question (Column 4).

2.1.2 Pre-study 2

Pre-study 2 explored whether first graders were familiar with the situations generated by the teachers in Pre-study 1, whether they considered these situations to be justice-relevant, and whether they named additional justice issues that were not mentioned by the teachers. A total of 19 first graders participated in the study. The children were divided into five groups and interviewed by group. The interview began with an informal talk about justice. The children were asked what they considered to be just and unjust. At the end of this talk, the interviewer summarized the discussion and stated that justice means treating everybody appropriately and giving everybody what they deserve.

Subsequently, the children were given descriptions of the situations from Table 1 and asked whether they were familiar with them. All children confirmed familiarity with each situation. Next, a specific event or behavior in each situation was described, for example: "If the teacher reads a story to you, can you choose the story?" The students were asked whether they considered the event or behavior to be justice-relevant ("Does this have anything to do with justice?") and if yes, whether they considered

Table 1 Results of Pre-study 1: Justice-relevant situations in the classroom

Justice form	Situation	Total number of situations generated by the teachers	Number of teachers who described the situation at least once
Retributive	1. Enforcing class rules	9	7
	2. Giving appropriate punishment	12	8
	3. Punishing the right person	10	8
Distributive	4. Evaluating task performance	14	6
	5. Grading achievements	29	10
	6. Allocating feedback	10	6
	7. Allocating popular tasks	6	3
	8. Allocating teacher's attention and help to children	17	9
	9. Allocating desired goods	7	6
	10. Allocating praise	12	8
	11. Allocating free choices (e.g., where to work or whom to work with) vs. restrictions	15	8
Interactional	12. Kind, respectful, and supportive treatment of child by teacher	10	8
	13. Kind, respectful, and supportive treatment of child by peers	10	6
	14. Appropriate and sensitive feedback by teacher	6	3
Procedural	15. Developing class rules and procedures for their enforcement	13	8
	16. Adaptive teaching (appropriate tasks, appropriate support)	16	9
	17. Explaining important information (e.g., teacher expectations, homework)	14	10

it to be just or unjust. The students rated the following situations as unjust: inappropriate punishment from the teacher, not getting help from the teacher when the child needs help, and not receiving praise after a child has worked hard. Other situations that the teachers named such as grading or adaptive learning were not perceived as justice-relevant by the children.

Finally, the students were asked whether they knew of additional situations in school that were justice-relevant. Two additional situations were mentioned by the majority of children: the opportunity to speak and to be listened to (Situation 18) and the amount of parental involvement in school activities (Situation 19).

2.2 Development of observer low inference and high inference instruments

As we explained earlier, some justice-relevant events and behaviors in the classroom are discrete, obvious, and can be observed directly, such as the feedback a child receives from the teacher (Situation 6). Other justice-relevant events and behaviors, such as adaptive teaching (Situation 16), are complex processes that cannot be easily decomposed into discrete observable elements. Rather, the elements of such complex processes have to be integrated mentally by observers and transformed into a justice rating. Therefore, we decided to develop both an observer low inference rating instrument and a high inference rating instrument. The observer low inference rating instrument was used by trained observers who observed the teacher and a group of up to five children over a period of 2 h in class. The high inference rating instrument was used by the observers and the teachers.

In order to come up with a set of (a) observable events and behaviors that could serve as the items for the observer low inference rating instrument and (b) rating items that could serve as the items for a high inference rating instrument, an expert group consisting of four teachers and two educational psychologists was established. The experts were given all of the justice-relevant situations described by the teachers in Pre-study 1 and the children in Pre-study 2. This material included descriptions of specific events and behaviors in these situations. The experts were asked to decide which justice-relevant events and behaviors could be observed, which ones could be rated, and which ones would be difficult or impossible to either observe or rate. All questions were discussed among the experts until a consensus was achieved. Subsequently, the experts were asked to propose items. The proposals were discussed in the group until a consensus was reached on how to phrase the items and response scales.

2.2.1 Observer low inference rating instrument

The observer low inference rating instrument consists of discrete events and behaviors such as being given the opportunity to speak. We call these discrete events and behaviors raw items. Some of these raw items are directly justice-relevant such as the number of times the teacher actively listens to a child. Other raw items are meaningful indicators of justice only after they are combined with other raw items. Consider the opportunity to speak (Situation 18) as an example. The opportunity to speak cannot be measured by the number of times a child is allowed to speak (Raw item 9) because the child might be called on without having a need to speak at that moment.

In addition, the child might express a need to speak by indicating this need (Raw item 8) but not be given the opportunity to speak. Therefore, we did not only count how often the child was called on but also how often the child was called on without having indicated a need to speak (Raw item 10). Dividing the number of times the child was called on (Raw item 9) minus the number of times he or she was called on without indicating a need to speak (Raw item 10) by the number of times he or she indicated a need to speak (Raw item 8) told us how often the child's need to speak was met. Items for which several raw items are combined into a meaningful justice indicator are called complex items.

The final observer low inference rating instrument was composed of the following 17 observable raw items. Table 2 presents the justice form (Column 1), the raw item(s) (Column 2), the complex item (Column 3), the situation (Column 4), and the justice-relevant need that is addressed by the situation (Column 5) forms.

The following complex items were composed of raw items:

Distributive justice:

$$18. \text{Opportunity to speak} = \frac{(1 \text{ child is allowed to speak} - 12 \text{ child is asked to speak without indicating})}{10 \text{ child indicates a need to speak}}$$

(Need: Child has a need to learn. Therefore, he or she needs opportunities to speak to improve his or her language skills and to feel valued by being allocated speaking time).

$$19. \text{Feedback} = \frac{8 \text{ no feedback, though desired}}{9 \text{ child indicates a need for feedback}}$$

(Need: Child has a need to learn. He or she requires feedback in order to improve.)

Procedural justice:

$$20. \text{Enforcing class rules 1} = \frac{3 \text{ teacher ignores the child's disruption}}{1 \text{ child disrupts the lesson}}$$

$$21. \text{Enforcing class rules 2} = - \frac{2 \text{ teacher accepts the child's disruption}}{1 \text{ child disrupts the lesson}}$$

(Need: Child has a need to improve social skills. He or she should learn the consequences of not following the class rules.)

2.2.2 High inference rating instruments

A total of 12 rating items were developed by the expert group. The wording of these items and the justice form they represent are given in Table 3. All items were combined with a four point rating scale (1/do not agree at all ... 4/completely agree) and an additional "not applicable" response option.

Table 2 Items included in the low inference rating instrument

Justice form	Raw item	Complex item	Situation	Need
P	1. Child disrupts the lesson 2. Teacher accepts the child's disruption 3. Teacher ignores the child's disruption	Items 1–3 are combined into a complex item	1	
D	4. Feedback concerning performance		6	Child has a need to learn and requires feedback to improve performance
D	5. Feedback concerning social behavior		6	Child has a need to improve social skills and requires feedback
D	6. Feedback concerning orderliness		6	Child has a need to learn and requires feedback to improve orderliness and organizational skills
D	7. Student indicates a need for feedback 8. Does not receive feedback even though child asked for feedback	Items 7 and 8 are combined into a complex item	6	
D	9. Child receives praise		10	Child has a need for positive reinforcement as a sign of recognition and also to develop his or her learning motivation
D	10. Child indicates a need to speak 11. Child is allowed to speak 12. Child is asked to speak without the child indicating the need	Items 10–12 are combined into a complex item	18	
I	13. Child is listened to actively		12	Child has a need for social affiliation. Being listened to is a sign of respect
I	14. Teacher makes dismissive remark about child		12	Child has a need for social affiliation
I	15. Child receives help		13	Child has a need for social affiliation. Therefore, he or she has to receive help and support from his or her peers
I	16. Child is denied help		13	Child has a need for social affiliation. Need is not being met if he or she is denied help
I	17. Child is disturbed by a peer		13	Child has a need to learn and for social affiliation. If he or she is disturbed by a peer, this need is not being met

P procedural justice, *I* interactional justice, *D* distributive justice

2.3 Main Study

2.3.1 Participants

Students A total of 208 students from seven primary schools participated in the study. Altogether, 15 classes took part. Four schools were located in rural areas, while the other three were located in an urban environment. 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 himself or herself had to agree to participate. The classes differed in size: The smallest class consisted of only 14 students, while the largest was composed of 30 students. In each class, at least eight and a maximum of 25 children took part in the study. 110 of the students were female (52.9%), the mean age of the students was 7.3 years ($SD=0.87$ years), and the students attended grades one to four. Ninety-two of the students were in a regular first-year primary school class. The other 116 students attended classes with mixed age groups in which children from grades one to four were taught together. They attended the following grade levels: $N=145$ were in grade one, $N=35$ in grade two, $N=6$ in grade three, and $N=17$ in grade four (missing: $N=5$).

Observers The observers taking part in the study were either students of psychology or prospective teachers. Altogether, 8 observers participated. All of them were female with an age range of 20–38 years.

Teachers 15 teachers (14 females; mean teaching experience: 9.6 years) took part in the study. All of them were primary school teachers and the classroom teacher in the participating classes.

2.3.2 Instruments

Observer low inference rating instrument Eight observers were trained to use the observer low inference rating instrument as well as the high inference rating instrument. The training included explanations of the single observation items, practice observations, and a debriefing to calibrate high inference ratings across observers. First, the observers read each item and a short explanation. Then they watched videos of preschoolers being taught a language topic by their teacher. Afterwards, they were given instructions about how to use the high inference rating instrument and about the meaning of each rating item. The observers then discussed examples of which situations to rate and when. To test how well the observer low and high inference ratings agreed after the training, we computed intra-class correlations for the raw observer low inference rating items as well as for the observer high inference rating items. Observers had to apply the observer low inference and high inference rating instruments to two videos that both showed a teacher interacting with four children. The duration of the video was 12 min. The videos were recorded for a German study

that evaluated the quality of language training with preschoolers. To analyze these data, we checked for consistency instead of absolute agreement. The items showed different Intra-class correlation scores (ICC). ICC scores also depended on the homogeneity of the group, and for some items, there was only a little variance. Also, some items were not observable in the presented videos at all. The items for which we could compute an ICC had an average ICC of 0.75.

Observer and teacher high inference rating instrument The high inference rating instrument items were rated on a four point rating scale (1 = do not agree at all to 4 = completely agree) and an additional not applicable response option. The 12 items covered the four justice forms. Based on the factor structure of an exploratory factor analysis the items were combined into indices by taking the mean of the items comprising each index. Table 7 and 9 present the results of the high inference rating indices. The high inference rating instrument for the teachers was the same as the observers' high inference rating instrument, with the exception of wording and phrasing. This was changed in order to present the questions concerning 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 on which the observation occurred.

Student high inference rating instrument Earlier, we argued on the basis of developmental theories and findings that children at the age of our sample are not able to provide differentiated justice ratings with sufficient validity. The results of our interviews show that students would not be able to provide detailed (low inference) ratings of justice but only global judgments (high inference ratings). For this reason, we decided to develop a simple rating instrument consisting of a general rating of the teacher's fairness and the fairness of the classmates as broad indicators of the child's sense of classroom justice: (a) Is your teacher fair (if the child did not understand the words "fair" or "just", then the observer explained the question in the following way: Does he/she treat you and your peers right? Does he/she treat everyone as he or she deserves to be treated)? (b) Are your peers fair? Both items were administered as part of an interview with each child after class and answered by the child on a four-point rating scale (1 = not true at all to 4 = completely true). The two items ($\alpha=0.64$) were aggregated, and their average was computed as a scale score ($M=3.11$; $SD=0.71$).

2.3.3 Procedure

Data collection in each class was conducted in the morning of days in February 2014. This point in time was chosen for data collection because the first graders were already familiar with their school environment, the rituals, and the rules, but they were still at the beginning of their school experience.

Depending on the size of the class, two to five observers were present in the classroom. Children were assigned randomly to the observers; every child was observed

Table 3 Items included in the high inference rating instruments

Justice form	Rating Item	Situations	Explanation
R	1. Teacher reacts to students' disturbances appropriately	2	Child has a need to learn to become a part of society. Therefore, it is necessary to be corrected if a behavior disrupts the class
P and D	2. Student is given enough time to think	8, 16	Child has a need to learn. The teacher has to develop and implement procedures (P) in which every student is given ample opportunity to think. There is only a restricted amount of time that is allocated to a topic during lecture time (D)
P	3. Student may finish work that was begun	11, 16	Child has a need to learn; this is not met if a child is stopped during his or her learning process. There is a mechanism developed by the institution to allocate learning opportunities in a certain way
P	4. Student is made to finish his or her work	11, 16	Child has a need to learn. If a slow student is not encouraged to finish a task in order to continue the lecture, then there is procedural injustice
P	5. Learning arrangements enable individual learning	16	Child has a need for academic achievements. Adaptive learning ensures that every child is challenged and supported
I	6. Student is listened to actively	12	Child has a need for social approval. It is a sign of respect if the teacher shows that he or she cares about what the child is saying
I	7. Tone between teacher and student is respectful	12	Child has a need for social approval. This need is being met if the teacher interacts with a respectful tone
I	8. Teacher's tone is appropriate	12	Child has a need to improve his or her social skills. The teacher as a model offers an appropriate tone that the child can use as feedback for his/her behavior
I	9. Praise and criticism are constructive	14	Child has a need to improve social and academic skills. Constructive feedback supports the child and his or her learning
I	10. Students reacts positively to feedback	14	Child has a need to improve social and academic skills. The child's reaction is an indicator that the teacher gave sensitive feedback.
D	11. Student gets praise for good performance	10	Child has a need to learn. Therefore, he or she requires positive reinforcement
D	12. Student has ample opportunities to speak up	18	Child has a need to learn. Therefore, he or she needs to be given enough time to speak to be able to get feedback and to improve

R retributive justice, *P* procedural justice, *I* interactional justice, *D* distributive justice

and each observer observed up to five children at the same time for a period of 120 min. During this time, both the students and the teacher were present in the classroom.

2.3.4 Data analysis

For data analysis, we used the statistical program "R" and the packages *psych*, *parallel*, and *MASS* as well as the program *SPSS 22*.

3 Results

3.1 Frequency of justice-related incidents and behaviors

Figure 1 gives the mean frequency scores (averaged across observers) of the raw items from the observer low inference rating instrument during 120 min of lecture time. As can be seen, the frequency of justice-relevant events and behaviors varied greatly. In average, a child raised his/her hand to be allowed to speak nine times but was allowed to speak only four times. Over the course of one school morning, a child was actively listened to only one to two times. Dismissive teacher remarks about a student were not observed at all.

3.2 Factor structure of the observer low inference rating items

The observer low inference rating instrument should allow us to measure procedural justice, distributive justice and interactional justice. Because different forms of justice and their indicators may or may not co-occur, we had no strong assumption about the factor structure of the items. Therefore, an exploratory factor analysis (EFA) seemed more appropriate as compared to a confirmatory factor analysis (CFA) of the items. Given that our data have a multi-level structure with children (level 1) being nested

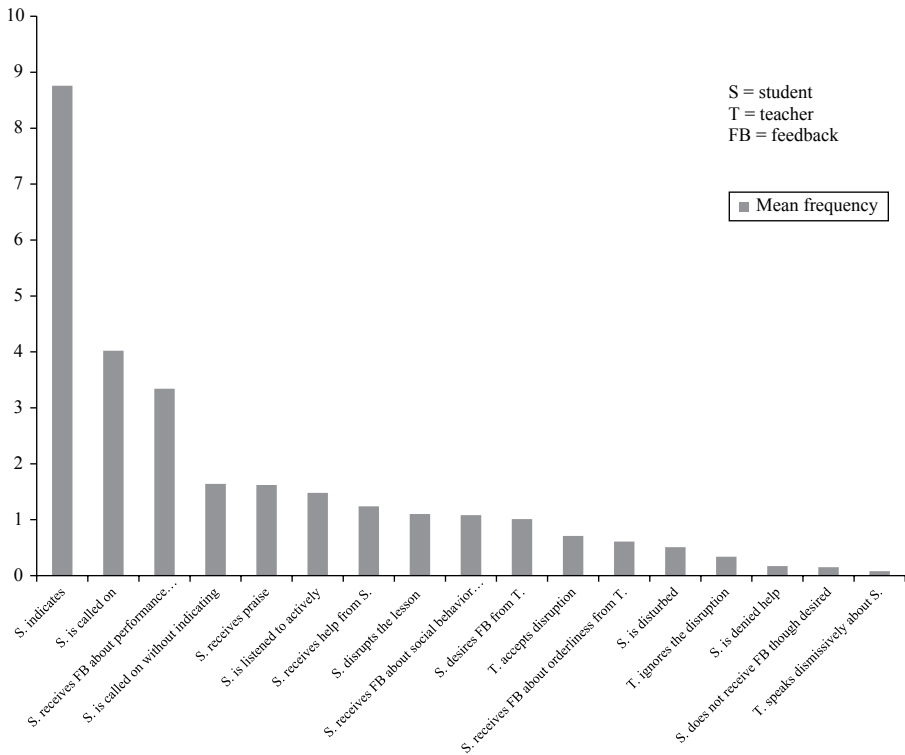


Fig. 1 Mean frequencies of the raw observer low inference rating items

in classes (level 2), correlations among items reflect common sources of variance at both levels. In order to separate both types of common sources of variance, the EFA was performed twice using the original items and residualized items. Residualized items were obtained via partialing out the class effect. This was achieved by dummy coding classes and regressing each item on all dummy variables. This procedure was conducted for all factor analyses as well as for the correlations between the justice instruments.

Initially, the factorability of all original and residualized items was examined. Bartlett's Tests of Sphericity and Kaiser-Meyer-Olkin tests showed that we could proceed with a factor analysis for original and residualized items even though the items are not metrically scaled and do not follow a normal distribution. The communalities were all above 0.30, further confirming that each item shared some common variance with other items. Given these overall indicators, factor analysis was conducted with all 13 items both as original items and as residualized items. Because some of the items are strongly skewed, we also computed polychoric correlations and performed the factor analyses with polychoric correlations. The results were virtually identical with those for the product moment correlations. Therefore, we report only the factor analytic results for the product moment correlations.

Parallel analysis suggested five common factors but inspection of the scree plot showed that the eigenvalue of the fifth factor did not meet the scree-criterion. This was true for both types of items (original, residualized). Analyses with five and with four factors were therefore conducted. Finally, the four factor model that explained 53 % of the residualized item variance (53 % of the original item variance) was preferred and rotated to simple structure because of the insufficient number of primary loadings and difficulty of interpreting the fourth factor and the fifth factor of the five factor solution. There was little difference in loadings between the varimax and the oblimin solutions, therefore only the results for the varimax solution will be reported.

The factor loadings of the four factor varimax solution are presented in Table 4 both for the residualized items and the original items. Loadings of the original items are presented in parentheses. Only loadings >0.20 are depicted. Item 16 (Child is denied help by a peer) failed to meet the criterion of having a primary factor loading of 0.35 or above (with the original as well as with the residualized items) and was therefore dropped from further analysis.

Based on the factor loadings, items were combined into composite scores. Composite scores were created for each of the factors by averaging the items that had their primary loadings on the factor to be measured. Items with negative loadings were recoded before building the composite score. We call the composites indices because their internal consistencies were rather low in some cases. Even though a low internal consistency does not necessarily imply a low reliability of a measure that contains heterogeneous items reflecting a broad or complex construct, calling the composites indices seems more appropriate than calling them scales.

Table 5 gives the names of the observer low inference rating indices (Column 1), the items that make up the indices (Column 2), the number of cases for which valid scores could be computed (Column 3), the means and the standard deviations of the indices with the original items (Columns 4 and 5), as well as their Cronbach's Alphas

Table 4 Factor loadings of the residualized observer low inference rating items (controlling for common class effects) and of the original low inference rating items

Observation items	Factors			
	1: performance feedback	2: enforcing class rules	3: respectful interactions	4: acceptance of the child
4. Feedback concerning performance	0.81 (0.81)			
5. Feedback concerning social behavior				-0.73 (-0.70)
6. Feedback concerning orderliness	0.40 (0.41)			-0.46 (-0.46)
9. Child receives praise	0.80 (0.80)			
13. Child is listened to actively	0.27 (0.28)		0.77 (0.77)	
14. Teacher makes dismissive remark about child	-0.27 (-0.26)		-0.24 (-0.24)	0.46 (0.46)
15. Child receives help		-0.21 (-0.24)	0.45 (0.44)	
16. Child is denied help			-0.71 (-0.71)	
17. Child is disturbed by a peer			-0.35 (-0.36)	0.31 (0.30)
18. Opportunity to speak				0.57 (0.57)
19. Feedback (recoded)	0.55 (0.53)	0.33 (0.37)		0.30 (0.28)
20. Enforcing class rules 1		0.93 (0.93)		
21. Enforcing class rules 2 (recoded)		0.91 (0.91)		

Loadings of the original low inference rating items are presented in parentheses

(Column 6). Alpha values of the residualized items appear first and Alphas for the original items are given next and in parentheses.

3.3 Factor structure of the high inference rating items

3.3.1 Observer high inference ratings

Next, an exploratory factor analysis was conducted on the items from the observer high inference ratings. Parallel analysis suggested five common factors, but the last two factors were not identified by the scree plot. This was true for both types of items (original, residualized). Therefore, the three-factor solution was chosen. The three-factor model explained 50% of the variance in the residualized items (46% in the original items). The factor loading matrix from the varimax solution is presented in Table 6. Again, loadings are given for both the original items and the residualized items. Loadings of the original items are presented in parentheses. Only loadings >0.20 are given. Item 9 (Praise and criticism are constructive) and Item 11 (Student gets praise for good performance) of the residualized and of the original items failed to meet the simple structure criterion of having no cross-loadings of 0.30 or above and were therefore not included in the indices.

Table 5 Descriptive statistics for the four observer low inference rating indices

Observer low inference rating indices based on factor analysis	Items	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Alpha</i>
LIR_O_Index1 supportive performance feedback	4. Feedback on performance 9. Child receives praise 19. Feedback (recoded)	207	2.07	1.68	0.61 (0.61)
LIR_O_Index2 enforcing class rules	20. Enforcing class rules 1 21. Enforcing class rules 2 (recoded)	87	-0.19	0.34	0.87 (0.87)
LIR_OLIR_Index3 respectful interactions	13. Child is listened to actively 15. Child receives help 16. Child is denied help	207	0.98	1.12	0.44 (0.43)
LIR_OLIR_Index4 acceptance of the child	5. Feedback concerning social behavior (recoded) 6. Feedback concerning orderliness (recoded) 14. Teacher makes dismissive remark about child 18. Opportunity to speak	207	-0.39	0.72	0.42 (0.41)

LIR_O Low Inference Ratings provided by Observers, *N* number of valid cases, *M* mean, *SD* standard deviation

Alphas in parentheses refer to indices built from the original items

The indices were built according to the primary factor loadings. Composite scores were created for each of the factors by averaging the mean of the items that had their primary loadings on the factor to be measured. Higher scores indicated a higher justice rating. Item 4 was recoded prior to building the index it belonged on. Table 7 gives the names of the observer high inference rating indices (Column 1), the items that made up the indices (Column 2), the number of valid cases (Column 3), the means and the standard deviations of the indices with the original items (Columns 4 and 5), as well as their Cronbach's alphas (Column 6). Alpha values of the residualized items appear first and Alphas for the original items are given next and in parentheses.

3.3.2 High inference teacher rating

The following step was to submit the high inference teacher ratings to an exploratory factor analysis. Parallel analysis suggested five common factors but, similar to the observer rating factors, the last two factors were not suggested by the scree criterion. Therefore, the three-factor solution was chosen. It explained 50% of the variance in the residualized items (57% variance in the original items). Orthogonal (varimax) and oblique (promax) solutions differed only marginally. The factor loading matrix from the varimax solution is presented in Table 8. Only loadings >0.20 are given. The loading patterns of the residualized items differed from the loading structure of the original items. The residualized items 1, 6, 7, and 12 failed the simple structure criterion of having no cross-loadings of 0.30 or above. Yet because their primary loadings were so high, it seemed justifiable to include these items in their primary indices. The original items 6, 9, and 12 had to be eliminated from further analyses because they either failed to meet the criterion of having a primary loading of 0.35 or above

Table 6 Factor loadings for the residualized observer high inference rating items (controlling for common class effects) and for the original observer high inference rating items

Observer rating items	Factors		
	1: adaptive learning settings	2: respectful teacher interaction	3: appropriateness of praise and criticism
1. Teacher reacts to students' disturbances appropriately		(0.23)	0.72 (0.72)
2. Student is given enough time to think	0.78 (0.77)		
3. Student may finish work that was begun	0.51 (0.52)	-0.44 (-0.40)	
4. Student is made to finish his or her work		-0.47 (-0.38)	
5. Learning arrangements enable individual learning	0.59 (0.60)		
6. Student is listened to actively	0.67 (0.66)	0.23 (0.25)	
7. Tone between teacher and student is respectful		0.50 (0.52)	
8. Teacher's tone is appropriate		0.55 (0.55)	
9. Praise and criticism are constructive	0.32 (0.29)	0.48 (0.50)	0.52 (0.52)
10. Student reacts positively to feedback			0.71 (0.71)
11. Student is praised for a good performance	0.45 (0.46)	0.23 (0.24)	0.46 (0.46)
12. Student is given enough opportunities to speak up	0.63 (0.62)	0.29 (0.32)	-0.34 (-0.34)

Loadings of the original low inference rating items are presented in parentheses

or because they failed the simple structure criterion of having no cross-loadings of 0.30 or above.

The indices were built according to the factor loadings. Composite scores were created for each of the factors by averaging the mean of the items that had their primary loadings on the factor to be measured. Higher scores indicated a higher justice rating. Table 9 gives the names of the high inference teacher rating indices (Column 1), the items that make up the indices of the residualized items and the indices of the original items, presented in parentheses (Column 2), the means and the standard deviations of the indices (Columns 3 and 4), and their respective Cronbach's alphas (Column 5).

3.4 Convergence between ratings

3.4.1 Observer low inference ratings and observer high inference ratings

In order to analyze the relations between the observer low inference ratings and the observer high inference ratings, spearman rank order correlations between the indices were computed. Significant correlations occurred between LIR_Index4 (acceptance of the child) with HIR_O_Index1 (adaptive learning setting) and HIR_O_Index3 (appropriate feedback). LIR_Index1 (performance feedback) and LIR_Index2 (enforcing class rules) had no significant correlations with any of the observer high

Table 7 Descriptive statistics for the observer high inference rating indices

Observer high inference rating indices based on factor analysis	Items	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Alpha</i>
HIR_O_Index1 adaptive learning settings	2. Student is given enough time to think 3. Student may finish work that was begun 5. Learning arrangements enable individual learning 12. Student is given enough opportunities to speak up	194	2.47	0.56	0.60 (0.59)
HIR_O_Index2 respectful teacher interaction	4. Student is made to finish his or her work (recoded) 7. Tone between teacher and student is respectful 8. Teacher's tone is appropriate	181	1.68	1.76	0.44 (0.56)
HIR_O_Index3 appropriateness of praise and criticism	1. Teacher reacts to students' disturbances appropriately 10. Student reacts positively to feedback	142	2.61	0.65	0.41 (0.43)

HIR_O High Inference Ratings provided by Observers, *N* number of valid cases, *M* mean, *SD* standard deviation

Alphas in parentheses refer to indices built from the original items

inference rating indices. This was true for the indices with the residualized items as well as for the indices with the original items. In addition to these correlations, there were significant correlations between HIR_O_Index1 (adaptive learning settings) and LIR_Index3 (respectful interactions) and LIR_Index4 (acceptance of the child) with the residualized items only. The number of cases for the correlations with LIR_Index2 is low because of the complex items it contains. These items are built of ratios and if a child for example never did "disrupt the lesson" then the divisor was zero and no ratio could be computed (Table 10).

3.4.2 Observer low inference ratings and teacher high inference ratings

As shown in Table 11, there are obvious differences between the analyses with residualized items and the ones with the original items. Taking the class level into consideration improved the relations between the observer low inference ratings and the teacher high inference ratings. The original items of the HIR_T_Index1 (appropriateness of criticism) correlated significantly with LIR_O_Index1 (performance feedback) and LIR_O_Index4 (acceptance of the child). This was also true for the original items. HIR_T_Index2 exhibited as the original items a significant correlation with LIR_O_Index4 (acceptance of the child). In addition, the residualized items of HIR_T_Index2 (adaptive learning settings) correlated also with LIR_O_Index1 (performance feedback) and with LIR_O_Index2 (enforcing class rules). Original as well as residualized items of HIR_T_Index3 (ensuring learning opportunities) exhibited correlations with LIR_O_Index2 (enforcing class rules). Residualized items of HIR_T_Index3 (ensuring learning opportunities) also correlated negatively with LIR_O_Index 3 (respectful interactions). No other correlations were significant. One teacher failed to hand in her ratings and the teachers made use of the additional

Table 8 Factor loadings for the residualized teacher high inference rating items (controlling for common class effects) and for the original teacher high inference rating items

High inference teacher rating items	Factors		
	1: adaptive learning settings	2: respectful teacher interactions	3: ensuring learning opportunities
1. Teacher reacts to students' disturbances appropriately	0.69 (0.69)	0.43	0.49 (0.49)
2. Student is given enough time to think	(0.21)	0.21	0.60 (0.65)
3. Student may finish work that was begun	0.77 (0.80)		
4. Student is made to finish his or her work			0.79 (0.82)
5. Learning arrangements enable individual learning	0.78 (0.73)	(0.22)	
6. Student is listened to actively	(0.37)	0.59 (0.30)	
7. Tone between teacher and student is respectful	0.21	(0.47)	0.75 (0.60)
8. Teacher's tone is appropriate	0.59 (0.36)	0.38 (0.65)	
9. Praise and criticism are constructive	0.30	0.84	
10. Student reacts positively to feedback		0.54 (0.66)	
11. Student is praised for a good performance		0.69 (0.52)	
12. Student is given enough opportunities to speak up	0.64 (0.56)	0.43 (0.57)	0.21

Loadings of the original low inference rating items are presented in parentheses

answer option (*not applicable*). Therefore, the case numbers used for our correlations vary between 144 and 146.

3.4.3 Observer low inference ratings and student high inference ratings

The student high inference ratings served as a fourth source of information about classroom justice. The residualized student high inference rating index had no significant correlations with the residualized observer low inference rating indices. The original student high inference rating index was correlated only with the original LIR_O_Index4 (acceptance of the child), and this correlation was small despite being significant: $r(184) = .18, p < .01$.

3.4.4 Observer high inference ratings and teacher high inference ratings

Most of the correlations between the teacher high inference ratings and the observer high inference ratings became larger when the common class effects was controlled. Specifically, HIR_O_Index3 (respectful interaction) correlated substantially with all three teacher high inference rating indices. In addition, there was a small correlation between HIR_O_Index2 (appropriate feedback) and HIR_T_Index3 (ensuring learn-

Table 9 Descriptive statistics for the teacher high inference rating indices

Teacher high inference rating indices based on factor analysis	Items	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Alpha</i>
HIR_T_Index1 appropriateness of praise and criticism	6. Student is listened to actively 9. Praise and criticism are constructive 10. Students reacts positively to feedback 11. Student gets praise for good Performance (original items in the index: 8, 9, 10, 11)	145	2.68	0.53	0.61 (0.67)
HIR_T_Index2 adaptive learning settings	1. Teacher reacts to students' disturbances appropriately 3. Student may finish work that was begun 5. Learning arrangements enable individual learning 8. Teacher's tone is appropriate 12. Student has ample opportunities to speak up (original items in the index: 1, 3, 5)	147	2.76	0.41	0.61 (0.67)
HIR_T_Index3 ensuring learning opportunities (non residualised items in the index: 2, 4, 7)	2. Student is given enough time to think 4. Student is made to finish his or her work (recoded) 7. Tone between teacher and student is respectful (original items in the index: 2, 4, 7) 10. Student reacts positively to feedback	146	2.36	0.61	0.61 (0.42)

HIR_T High Inference Ratings provided by Teachers, *N* number of valid cases, *M* mean, *SD* standard deviation

Alphas in parentheses refer to indices built from the original items

ing opportunities). Without controlling for the common class effect, the observer high inference rating indices and the teacher high inference rating indices were mostly independent. The only significant correlation was between HIR_O_Index1 (adaptive learning settings) and HIR_T_Index2 (adaptive learning settings) (Table 12).

4 Discussion

To our knowledge, we have presented the first low inference rating instrument for measuring justice in the classroom. In two pre-studies, we extracted justice-relevant situations and developed observable indicators for these situations. In the main study, items from the pre-studies served to observe 208 primary school students with regard to justice related events during class. In addition to the low inference ratings for observers, high inference rating instruments for observers, teachers, and students were employed for examining the convergence between low and high inference ratings of classroom justice across different judges (observers, teachers, and students).

Table 10 Correlations between the observer low inference ratings and the observer high inference ratings

Observer high inference ratings		Observer low inference ratings			
		LIR_O_Index1 performance feedback	LIR_O_Index2 enforcing class rules	LIR_O_Index3 respectful interactions	LIR_O_Index4 acceptance of the child
HIR_O_Index1 adaptive learning settings	<i>r</i>	.04 (.01)	-.01 (-.10)	.14* (.09)	.15* (.12)
	<i>p</i>	.55 (.91)	.91 (.37)	.03 (.19)	.03 (.07)
	<i>N</i>	207	87	186	207
HIR_O_Index2 respectful interaction	<i>r</i>	.02 (-.08)	-.10 (-.10)	.10 (-.02)	.19** (.26**)
	<i>p</i>	.80 (.29)	.39 (.40)	.15 (.75)	.00 (.00)
	<i>N</i>	198	87	207	198
HIR_O_Index3 appropriate feedback	<i>r</i>	.03 (.02)	-.11 (-.11)	.06 (-.05)	.23** (.16*)
	<i>p</i>	.68 (.73)	.31 (.32)	.41 (.52)	.00 (.02)
	<i>N</i>	207	85	198	207

r spearman correlation, *N* number of valid cases, *LIR_O* Low Inference Ratings provided by Observers, *HIR_O* High Inference Ratings provided by Observers, Values in parentheses refer the values of the original items

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

Table 11 Correlations between the observer low inference ratings and the teacher high inference ratings

Teacher high inference ratings		Observer low inference ratings			
		LIR_O_Index1 performance feedback	LIR_O_Index2 enforcing class rules	LIR_O_Index3 respectful interactions	LIR_O_Index4 acceptance of the child
HIR_T_Index1 appropriateness of criticism	<i>r</i>	-.21** (-.17*)	.14 (.08)	-.07 (-.06)	.28** (.25**)
	<i>p</i>	.01 (.04)	.27 (.51)	.39 (.46)	.00
	<i>N</i>	144	69	144	144
HIR_T_Index2 adaptive learning settings	<i>r</i>	-.29** (-.14)	.37** (.21)	-.03 (-.06)	.40** (.23**)
	<i>p</i>	.00 (.08)	.00 (.08)	.74 (.51)	.00 (.01)
	<i>N</i>	146	69	146	146
HIR_T_Index3 ensuring learning opportunities	<i>r</i>	-.03 (.06)	.31** (.25*)	-.17* (-.01)	.06 (-.12)
	<i>p</i>	.73 (.48)	.01 (.04)	.05 (.88)	.50 (.10)
	<i>N</i>	145	69	146	145

r spearman correlation, *N* number of valid cases, *LIR_O* Low Inference Ratings provided by Observers, *HIR_T* High Inference Ratings provided by Teachers, Values in parentheses refer the values of the original items

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

4.1 Objectivity, factorial structure, and reliability of the rating instruments

4.1.1 Observer low inference ratings

To examine the objectivity of the low inference ratings, we had eight observers use the instrument for the same children and the same situations in a classroom. We esti-

Table 12 Correlations between the observer high inference ratings and the teacher high inference ratings

Teacher high inference ratings		Observer high inference ratings		
		HIR_O_Index1 adaptive learning settings	HIR_O_Index2 appropriate feedback	HIR_O_Index3 re- spectful interaction
HIR_T_Index1 respectful teacher interactions	<i>r</i>	-.11 (-.07)	-.01 (.04)	.38** (-.14)
	<i>p</i>	.20 (.38)	.93 (.68)	.00 (.10)
	<i>N</i>	145	139	145
HIR_T_Index2 adap- tive learning settings	<i>r</i>	.15 (.24**)	-.08 (-.05)	.56** (-.04)
	<i>p</i>	.08 (.00)	.32 (.59)	.00 (.67)
	<i>N</i>	147	141	147
HIR_T_Index3 ensuring learning opportunities	<i>r</i>	-.03 (.08)	.17* (.14)	.32** (.08)
	<i>p</i>	.97 (.32)	.04 (.11)	.00 (.38)
	<i>N</i>	140	140	146

r correlation, *N* number of valid cases, *HIR_O* High Inference Ratings provided by Observers, *LIR_O* Low Inference Ratings provided by Observers, Values in parentheses refer the values of the original items

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

mated the interrater agreement using the intraclass correlation and found an average ICC of 0.75. This result suggests a satisfactory amount of objectivity.

The dimensional structure of the low inference ratings was determined using exploratory factor analysis. Four common factors were extracted and these were independent on whether or not the class effect was controlled for or not and independent on whether product moment correlations or polychoric correlations were used. The extracted factors reflect (1) performance feedback, (2) enforcing class rules, (3) respectful interactions, and (4) accepting the child and letting the child act.

Based on the factor loadings pattern, we built four indices by averaging items with primary loadings on a factor. We built two kinds of indices, a first one using the original items and a second one using residualized items (controlling for common class effects). Most of the indices had modest alpha levels. This was due to the small number of items and to the heterogeneity of the items. It is important to note that item heterogeneity decreases reliability but may well increase validity because heterogeneous items can provide a broader representation of a complex construct, whereas homogeneous items often reflect only a small facet of a broad construct (Yousfi 2005a, b).

4.1.2 Observer and teacher high inference ratings

Observer and teacher high inference rating items were also factor analyzed. In both cases, a three-factor solution was accepted. The observer high inference rating factors reflect (1) adaptive learning settings, (2) respectful teacher interactions, and (3) appropriateness of praise and criticism. The teacher high inference rating factors reflect (1) adaptive learning settings, (2) respectful teacher interactions, and (3) ensuring learning opportunities. As for the observer low inference ratings, the factorial structure was independent of the type of correlation used (product moment vs. polychoric). The factorial structure of the observer high inference ratings remained

unaffected by controlling for class effects. However, the factorial structure of the teacher high inference ratings differed when using original items vs. residualized items. We will discuss common class effects and their implications in more detail in the next section of the discussion. Before we do so, we draw some general conclusions from the factor analytic results.

4.1.3 Conclusions from the factor analytic results

The factor analytic results we have reported suggest at least three important conclusions. First, the justice dimensions that can be discriminated depend on the inference level of the ratings and, to a smaller extent, on who performed the ratings. Although the dimensions derived from the three rating instruments are similar, they are not identical. More specifically, it seems that low inference ratings enable raters to generate more differentiated justice evaluations as compared to high inference ratings. Possibly, high inference ratings blur differences and nuances that are noticeable on the level of specific behaviors of the child, the teacher, and peers. Notably, however, one dimension of justice appeared invariantly across all ratings: respectful teacher interaction. We will come back to this result in the next section of our discussion.

Second, the factor analyses of the low inference ratings and the two high inference ratings clearly show that classroom justice cannot be structured primarily according to the forms of justice at issue (distributive, punitive, procedural, interactional). Rather, specific situations seem to trigger justice-related behavior and generate variability in justice ratings. This might at least partly reflect the fact that many behaviors can be mapped on several forms of justice. Feedback, for example, can be fair or unfair in different ways: It can be a matter of distributive justice because the teacher has to take time during the lecture to give feedback. It can also be an issue of interactional justice depending on how the teacher phrases feedback.

Third, the high similarity between oblique and orthogonal solutions suggests that, unlike in the organizational fairness domain (Colquitt et al. 2001), different kinds of fair vs. unfair treatment of children in primary school do not seem to co-occur. The low internal consistencies of some of the indices support this conclusion. This result is remarkable for two reasons. First, it fully supports our decision to perform exploratory instead of confirmatory factor analyses. Second, it suggests that teachers do not have a generalized tendency of treating their students more or less justly independent of the specific context and the kind of interaction that occurs.

4.2 Common class effects

The data of our main study had a multi-level structure with students nested in classes and classes being confounded with teachers. Because teachers may differ in their overall level of treating students fairly, class effects can be confounded with individual differences between students in fair treatment. Therefore, we compared correlations among justice ratings with and without controlling for class effects. It turned out that low and high inference ratings of observers and students remained unaffected by class membership. By contrast, individual differences between students in fair treatment depended on class membership when rated by teachers. Compared to observ-

ers and students, teachers rated their justice behavior as being more similar across students. In other words: Compared to student high inference ratings and observer ratings, teacher ratings underestimate within classes differences and overestimate between classes differences in justice.

Controlling for this bias resulted in substantial correlations between the teacher high inference ratings and the observer low inference ratings. In fact, these correlations were even higher on average than were the correlations between the observer low inference ratings and the observer high inference ratings even though the latter two ratings stemmed from the same observers. One possible explanation for this effect could be that the external observers managed to perform the low inference ratings quite well but had difficulties in providing complex high inference ratings. These difficulties may be due to limited knowledge about the children and about how teachers and peers typically treat them. Teachers on the other hand are experts of their class and therefore their high inference ratings might reflect closer the more objective low inference ratings of the observers. In line with this idea, all teacher high inference rating indices, once adjusted for the common class effect, exhibited moderate to high correlations with the observer high inference rating index 3 (respectful interaction). This index may be best suited for observer high inferences ratings of justice in cases when observers have only a limited number of time available for observing a large number of specific justice-related events. Because respectful interaction, a facet of interactional justice, is probably not bound to a specific event but rather a generalized manner in which a teacher treats a child, it can be observed more frequently and this might explain why this index converged best with the teachers' self-evaluation.

4.3 Convergence between ratings

In the previous sections, we have addressed possible explanations for the variable convergence among the justice ratings. We discussed why the teacher high inference ratings might converge more strongly with the observer low inference ratings than with the observer high inference ratings. We have also hinted at possible reasons for the low convergence between the low and high inference ratings of observers. Regarding the latter, we would like to add an explanation. It may not only be the case that observers have trouble integrating concrete observations into a more general justice judgment. It may also be true that concrete events considered justice relevant by observers that did not appear in the low inference rating instrument nevertheless affected high inference ratings.

The student high inference ratings showed no correlation with the observer low inference ratings. What could be the reason for this lack of convergence? It is reasonable to assume that especially young students are not yet able to rate classroom justice accurately. As we mentioned in the introduction, primary school students still struggle to give valid answers in interviews due to their limited vocabulary, their limited conceptual understanding, and because they have difficulties aggregating and abstracting information from single incidents (Piaget 1997; Biemer and Lyberg 2003; Goswami 2011). But even for older students who are able to assess complex constructs such as instructional quality, there is not necessarily a relation between their justice ratings and their teachers' high inference ratings. Eckstein and Noack (2014)

found that there was no correlation between teachers' and high school students' high inference rating of the fairness of the teacher. Therefore, it is possible that the student high inference ratings measure something that was not captured by observable indicators of classroom justice. If this is true, then the student high inference ratings and the observer low inference ratings both offer unique information about the complex concept of classroom justice. Clausen et al. (2002) concluded the same for another complex construct: instructional quality. There too, the correlation between different protagonists in school were very low. The lack of elaborate justice concepts of young children and their limited ability of integrating justice relevant incidents into a justice rating does not necessarily imply that they do not have a sense of justice. It also does not imply that experiences of injustice have no emotional, attitudinal, and behavioral consequences. Rather, how children perceive justice may differ from the way adults do without justice being less important for the former than for the latter.

Together, the limited overlap between the various measures of classroom justice implies that no single measure taps classroom justice comprehensively. This results from the occasion- and situation-specificity of justice-related behavior. It also means that justice is in the eye of the beholder, i.e., a subjective matter and difficult to measure objectively. Accordingly, our findings suggest that a comprehensive assessment of classroom justice requires a multi-method approach. Measuring classroom justice, as was attempted in previous research, with a single method and only one perspective of the parties that are involved in justice-relevant situations, is not possible.

4.4 Limitations

Next, we address some limitations of our research. First, detailed feedback based on an observation accuracy test could have helped to train the raters even more intensely in the use of the rating instrument. However, the good interrater consistency indicated that the observers sufficiently agreed in their low inference ratings. Second, it could have been beneficial to include more items in the low inference ratings instrument. Third, further specific items might have been useful for disambiguating the meaning of interactions. For example, item 4 (Feedback concerning performance) could have been combined with an item for the valence of feedback (positive, negative, or neutral). Last but not least, the factor structure of all instruments is suggestive but not compelling. Future research has to address these issues.

4.5 Conclusions and practical implications

Although the factorial structure of our instruments and the (lack) of overlap among them awaits replication, our research is an important step towards assessing justice in the classroom more comprehensively as compared to previous research. Previous studies relied mostly on single sources of information, thus ignoring levels of justice judgments and ignoring that perceptions of justice depend strongly on the role a person plays in a justice event. Overcoming these limitations requires multi-method and multi-perspective approaches. Given that justice matters to students and teachers alike and that social scientists and lay people know the strong effects of justice on social behavior, it seems relatively easy to commit participants for justice research in

schools. In fact the teachers who took part in our study indicated that rating classroom justice gave them insights into their own teaching, and they were very interested in the results of the low inference ratings in order to uncover the unjust aspects of their classrooms and to enhance classroom justice. Teachers try to treat their students in a fair way (Kanders 2000), and getting an outside perspective on classroom justice could help them to achieve more fairness in their classes.

References

- Adams, J. S. (1965). Inequity in social exchange. In L. Berkowitz (Ed.), *Advances in experimental social psychology: Vol. 2*. (pp. 267–99). New York: Academic Press.
- Aristotle. (1998). *The Nicomachean ethics*. Oxford: Oxford University Press.
- Au, W., Bigelow, B., & Karp, S. (2007). *Rethinking our classrooms*. Milwaukee: Rethinking Schools.
- Baumert, A., & Schmitt, M. (in press). Justice sensitivity. In C. Sabbagh & M. Schmitt (Eds.), *Handbook of social justice theory and research*. New York: Springer.
- 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.
- Burnett, P. C., & Mandel, V. (2010). Praise and Feedback in the Primary Classroom: Teachers' and Students' Perspectives. *Australian Journal of Educational & Developmental Psychology, 10*, 145–154.
- 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.
- Chory-Assad, R. M. (2002). Classroom justice: Perceptions of fairness as a predictor of student motivation, learning, and Aggression. *Communication Quarterly, 50*, 58–77.
- Schnabel, K., & Schröder, S. (2002). Konstrukte der Unterrichtsqualität im Expertenurteil. *Unterrichtswissenschaft, 30*(3), 246–260.
- Colquitt, J. A., Conlon, D. E., Wesson, M. J., Porter, C. O. L. H., & Ng, K. Y. (2001). Justice at the millennium: A meta-analytic review of 25 years of organizational justice research. *Journal of Applied Psychology, 86*(3), 425–445.
- Connell, R. W. (1993). *Schools & social justice*. Sidney: Pluto Press.
- Corden, R. (2000). *Literacy & learning through talk: Strategies for the primary classroom*. Buckingham: Open University Press.
- 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.
- Correia, I., Kamble, S. V., & Dalbert, C. (2009). Belief in a just world and well-being of bullies, victims and defenders: A study with Portuguese and Indian students. *Anxiety, Stress, & Coping, 22*(5), 497–508.
- Crosby, F. (1984). Relative deprivation in organizational settings. *Research in Organizational Behavior, 6*, 51–93.
- Dalbert, C. (2011). Warum die durch die Schüler und Schülerinnen individuell und subjektiv erlebte Gerechtigkeit des Lehrerhandelns wichtig ist [Why subjective experience of teacher justice is important]. *Zeitschrift für Pädagogische Psychologie, 25*(1), 5–18.
- Dalbert, C., & Stoerber, J. (2005). The belief in a just world and distress at school. *Social Psychology of Education, 8*(2), 123–135.
- Damon, W., & Killen, M. (1982). Peer interaction and the process of change in children's moral reasoning. *Merrill-Palmer Quarterly, 28*(3), 347–367.
- Deutsch, M. (1985). *Distributive justice: A social-psychological perspective*. New Haven: Yale University Press.
- Ditton, H. (2002). Lehrkräfte und Unterricht aus Schülersicht. Ergebnisse einer Untersuchung im Fach Mathematik. [Teachers and instruction from students' perspective. Results from a study on mathematics]. *Zeitschrift für Pädagogik, 48*(2), 262–286.
- Eckstein, K., & Noack, P. (2014). Students' democratic experiences in school: A multilevel analysis of social-emotional influences. *International Journal of Developmental Science, 8*(3), 105–114.

- Gollwitzer, M., & Van Prooijen, J.-W. (in press). Psychology of justice. In C. Sabbagh & M. Schmitt (Eds.), *Handbook of social justice theory and research*. New York: Springer.
- Goswami, U. (2011). A temporal sampling framework for developmental dyslexia. *Trends in Cognitive Sciences*, 15(1), 3–10.
- Gruehn, S. (2000). *Unterricht und schulisches Lernen: Schüler als Quellen der Unterrichtsbeschreibung [Instruction and learning in school: Students as sources of describing instruction]*. Münster: Waxmann.
- Harris, P. L. (1989). *Children and emotion: The development of psychological understanding*. Oxford: Basil Blackwell.
- Hattie, J. (2009). *Visible learning*. London: Routledge.
- Henderlong, J., & Lepper, M. R. (2002). The effects of praise on children's intrinsic motivation: A review and synthesis. *Psychological Bulletin*, 128(5), 774–795.
- Honkanen, M., Määttä, H., Hurtig, T., Ebeling, H., Taanila, A., & Koivumaa-Honkanen, H. (2014). Teachers' assessments of children's mental problems with respect to adolescents' subsequent self-reported mental health. *Journal of Adolescent Health*, 54(1), 81–87.
- 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. (in press). Distributive justice. In C. Sabbagh & M. Schmitt (Eds.), *Handbook of social justice theory and research*. New York: Springer.
- Kanders, M. (2000). IFS-Umfrage: Die Schule im Spiegel der öffentlichen Meinung. Ergebnisse der elften IFS-Repräsentativbefragung der bundesdeutschen Bevölkerung [IFS Survey: The school in public opinion. Results of the 11th IFS representative survey of the German population]. In H.-G. Rolff, K.-O. Bauer, K. Klemm, H. Pfeiffer, & R. Schulz-Zander (Eds.), *Jahrbuch der Schulentwicklung* (pp. 13–50). Weinheim: Juventa.
- Lerner, M. J. (1977). The justice motive: Some hypotheses as to its origins and forms. *Journal of Personality*, 45(1), 1–52.
- Lerner, M. J. (1980). *The belief in a just world*. Boston: Springer.
- 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.
- Liebig, S., & Sauer, C. (in press). Sociology of justice. In C. Sabbagh & M. Schmitt (Eds.), *Handbook of social justice theory and research*. New York: Springer.
- Lind, E. A., & Tyler, T. R. (1988). *The social psychology of procedural justice*. Berlin: Springer Science & Business Media.
- Mayring, P. (2010). *Qualitative Inhaltsanalyse Grundlagen und Techniken [Qualitative content analyses Basics and techniques]*. Weinheim: Beltz.
- Meyer, L. H., & Sanklecha, P. (in press). The political philosophy of justice. In C. Sabbagh & M. Schmitt (Eds.), *Handbook of social justice theory and research*. New York: Springer.
- Mikula, G., Petri, B., & Tanzer, N. (1990). What people regard as unjust: Types and structure of everyday experience of injustice. *European Journal of Social Psychology*, 20, 133–149.
- Montada, L., & Lerner, M. J. (Eds.). (1998). *Responses to victimizations and belief in a just world*. New York: Plenum Press.
- Osterman, K. F. (2000). Students' need for belonging in the school community. *Review of Educational Research*, 70(3), 323–367.
- 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.
- Petillon, H. (1993). *Das Sozialleben des Schulanfängers: Die Schule aus der Sicht des Kindes*. Weinheim: Beltz Psychologie-Verlag -Union.
- Piaget. (1997). *The moral judgment of the child*. New York: Free Press Paperbacks.
- Pretsch, J., Ehrhardt, N., Engl, L., Risch, B., Roth, J., Schumacher, S., & Schmitt, M. (in press). Injustice in school and students' emotions, well-being, and behavior: A longitudinal study. *Social Justice Research*.
- Rawls, J. (1971). *A theory of justice*. Cambridge: Belknap Press of Harvard University Press.
- 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. (in press). Justice in education. In C. Sabbagh & M. Schmitt (Eds.), *Handbook of social justice theory and research*. New York: Springer.

- 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.) (in press). *Handbook of social justice theory and research*. New York: Springer.
- Schwan, G. (2008). *Gerechtigkeit und Bildung für eine Politik der Chancen*. Vortrag auf einer Veranstaltung der Friedrich Ebert Stiftung in den Franckeischen Stiftungen zu Halle (Saale).
- Stroet, K., Opdenakker, M.-C., & Minnaert, A. (2013). Effects of need supportive teaching on early adolescents' motivation and engagement: A review of the literature. *Educational Research Review*, 9, 65–87.
- Susteck, H. (1996). Rituale in der Schule [Rituals in school]. *Pädagogische Welt*, 50(1), 34–38.
- Tata, J. (1999). Grade distributions, grading procedures, and students' evaluations of instructors: A justice perspective. *The Journal of Psychology*, 133(3), 263–271.
- Wenzel, M., & Okimoto, T. G. (in press). Retributive justice. In C. Sabbagh & M. Schmitt (Eds.), *Handbook of social justice theory and research*. New York: Springer.
- Wubbels, T., & Brekelmans, M. (2005). Two decades of research on teacher–student relationships in class. *International Journal of Educational Research*, 43(1), 6–24.
- Yousfi, S. (2005a). Mythen und Paradoxien der Klassischen Testtheorie (I): Testlänge und Gütekriterien [Myths and paradoxons of Classical Test Theory (I)]. *Diagnostica*, 51, 1–11.
- Yousfi, S. (2005b). Mythen und Paradoxien der Klassischen Testtheorie (II): Trennschärfe und Gütekriterien [Myths and paradoxons of Classical Test Theory (II)]. *Diagnostica*, 51, 55–66.