Student Attitudes Toward the Assessment Process:

An Empirical Analysis of the Formation and the Consequences of

Grading-related Justice Perceptions

Dissertation

zur Erlangung des Doktorgrades

der Wirtschafts- und Sozialwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

vorgelegt von

Roland Burger, M.A.

aus Böblingen

Tübingen

2016

Tag der mündlichen Prüfung:	21.04.2017
Dekan:	Professor Dr. rer. soc. Josef Schmid
1. Gutachter:	Professor Dr. rer. soc. Martin Groß
2. Gutachter:	Professor Dr. phil. Steffen Hillmert

Table of Contents

List of Tables

List of Figures

1 Introduction

Throughout all stages of education, students have to demonstrate their efforts and abilities in a steadily unfolding sequence of assessments. Assessment is such a fundamental element of the modern curriculum that at times, receiving a good grade threatens to overshadow other important goals (such as the acquisition of knowledge and skills). In response to this, a number of scholars has recently gone so far as to actually call for discarding the general practice of grading student performance altogether (Lynch and Hennessy 2015; McMorran et al. 2015; Tannock 2015). While some of the arguments brought forward in these studies might eventually induce a shift in how assessment is organized, abandoning research on the subject of grading would be a little hasty – grades still hold immense power over the learner, and will likely continue to do so in the foreseeable future.

A lot can be gathered from the information contained in a grade. Grades give students explicit, albeit superficial, feedback about whether their efforts were outstandingly good, severely disappointing, or anything in between. Outside the immediate context of education, grades play a decisive role in signaling the students' abilities to potential employers (Jones & Jackson 1990; Spence 1973). Given the general importance of grades, it is reasonable to assume that students do not only feel entitled to receive a high quality education, but also to have their efforts appreciated through appropriate grades.

The focus of this thesis is how *fair* the procedures that generate these grades appear from the perspective of university students. It should come as no surprise to anyone who has been awarded grades at some point in their educational career that assessment is often accompanied by justice-related concerns[1]. Some students might feel that they did not get the grade they deserve, others might be troubled by inconsistent grading standards, again others might bemoan a lack of transparency. Since students invest considerable efforts in the attainment of grades, an absence of fairness can be the source of substantial distress (Jasso and Resh 2002; Nesbit and Burton 2006; Sabbagh et al. 2006).

Questions of how individuals experience the fairness of distributions and procedures are the subject of *organizational justice research* (Cropanzano and Ambrose 2015).

---

[1] The terms "justice" and "fairness" will be used interchangeably throughout this thesis.

This field has generated an immense body of literature spanning several decades and covering an extensive range of justice-related issues. The central theoretical concepts on which this thesis is built are *procedural justice* (Leventhal 1980; Thibaut and Walker 1975) and *informational justice* (Greenberg 1993). Procedural justice has been attracting substantial interest since the concept's inception in the mid-1970s, so that at present, we have a fairly good understanding of what defines a fair procedure and of why it is important that procedures are fair. This is complemented by a sizable body of research targeting informational justice, which asks about the fairness of explanations that relate to a procedure (Colquitt et al. 2013).

But even though nobody will refute the notion that grading procedures should be fair, this subject has rarely caught the attention of researchers. Thus far, the overwhelming majority of organizational justice research has been focused on justice in the workplace, investigating the fairness of procedures such as performance appraisals, employee selection, layoffs, and conflict resolution (Bobocel and Gosse 2015). Compared to that, the number of studies that have attended to questions of procedural justice and informational justice in education is strikingly limited. This is evidenced by the fact that a recent review of research on justice in education by Resh and Sabbagh (2016) is exclusively focused on distributive justice, mentioning procedural justice only briefly at the very end. This leaves matters of fair procedures in education a largely uncharted territory – a situation this thesis intends to change.

In order to provide a comprehensive picture of grading-related justice perceptions, the subject is approached from two complementary perspectives. The first part of the research agenda investigates the *consequences* of perceived injustice while the *formation* of these perceptions is scrutinized in the second part. By combining research on the outcomes of perceived injustice with investigations that target the development of these feelings, this thesis aims to problematize issues that stem from unfair grading procedures while at the same time seeking practicable solutions for an improvement of justice perceptions in a total of three empirical studies.

First, there is a need to establish the practical relevance of fair grading procedures. It can be safely assumed that unfair procedures are viewed negatively by all students. Given the importance of grades for success in education, it should be in the students' best interest that their efforts are judged using fair procedures. However, the question is whether violations of procedural justice represent a mere nuisance or if there are more severe consequences. In order to address this issue, Study 1 makes an attempt to link

student perceptions of the fairness of grading procedures to their intention to abandon their studies before graduating, i.e. to drop out.

The second part of the research agenda shifts the attention from the consequences of the perceived fairness of grading procedures to the fairness perceptions themselves by making them the dependent variables. Thus far, the vast majority of justice research has been focused on the consequences of justice perceptions: What happens if individuals experience injustice? A central proposition of this thesis is that this is not enough. Knowing about the consequences of justice perceptions is imperative, but the knowledge thus derived is inconsequential unless we also find ways to alleviate feelings of injustice. This places the focus on the formation of justice perceptions. With regard to this, this thesis asks two basic questions. The first question concerns *why* observed procedures align with or deviate from the theoretical "ideal" while the second question addresses *how* justice perceptions are formed through the interplay of various situational characteristics.

With regard to the *why*, it is essential to establish what actually determines whether students perceive the grading process as fair. The majority of organizational justice research has treated justice perceptions as individual level phenomena. In case antecedents of justice perceptions are investigated, the focus is usually on individual-level factors such as psychological dispositions (Vermunt and Steensma 2016). A recent line of research has started to question the focus on the individual by acknowledging that justice perceptions are also a product of their context. The *justice climate* approach proposes that individuals who share a common environment also make similar experiences (Naumann and Bennett 2000). Because of the similarity of experiences and because individuals share these experiences with each other, justice perceptions within groups tend to exhibit similarities (Li and Cropanzano 2009). But regardless of the insight that context matters, research on justice climate has thus far failed to acknowledge an important element: the role of *structural characteristics* of the environmental context that affect the chance that a procedure satisfies justice-related expectations. The research in this thesis is located in a university setting, where students are enrolled in a wide range of different departments. The goal of Study 2 is to identify structural components of these departments that affect individual justice perceptions, thus bringing about between-department differences in justice climate.

Concerning the *how*, this thesis seeks to clarify the mechanisms through which individuals combine different characteristics of a situation and weigh them against each other to form justice judgments. As will be discussed later on in greater detail, justice

research usually holds the fairness of a procedure to the extent to which it adheres to a number of rules or criteria (Colquitt and Rodell 2015). However, empirical studies rarely acknowledge the possibility that in practice, these rules might not be equally important across different contexts and procedures. But as Leventhal (1980) has stated in his seminal attempt to define procedural justice, "[…] it is assumed that an individual applies procedural rules selectively and follows different rules at different times" (p. 32). Therefore, understanding the formation of feelings of injustice necessitates an exploration of the preferential structure through which justice criteria are mapped to justice judgments. Study 3 takes on this subject by investigating how students judge the fairness of the feedback they receive following a graded assignment.

The remainder of this thesis is structured as follows. The following chapter starts with a general introduction to the field of empirical justice research which leads us to the subject of this thesis: the fairness of procedures. Starting with an outline of why it is important that procedures are fair, we move on to the procedures used in grading assessments of student performance. Here, the state of research on justice in education will be discussed, which will identify issues of previous research on the theoretical as well as the empirical level. The main point here will be to establish a coherent framework for the investigation of justice perceptions in education. Next, the various criteria that underlie students' justice perceptions are introduced and their meaning in the context of the assessment process is established. The primary focal points are the extent to which students can get involved in the grading process, the extent to which the grading process delivers results that can be deemed valid, and the fairness of assessment-related feedback provided to the students. In turn, the research agenda of this thesis will be derived from this discussion. The various parts of this research agenda were at the center of attention in three empirical studies. The aims, theoretical background, and methodological approach of each of these studies are introduced in chapter 3. Chapters 4 to 6 contain the manuscripts of the studies. This is followed by a discussion of the results of each study as well as of their role in the advancement of the state of research in chapter 7. Even though the present work specifically targets the fairness of grading procedures, the theoretical and methodological framework to be developed in this thesis will provide a template for future research on the fairness of a broad spectrum of other distributive procedures, a discussion of which bookends this thesis.

2 Theoretical framework

Whereas the subjects of other areas of sociological inquiry such as social stratification and mobility are of immense importance for shaping individual life courses and thus society as a whole, it can be assumed that a substantial part of the general population rarely reflects on these matters in everyday situations. In contrast, justice is an integral part of everyday thinking that is not tied to any specific context. Even people completely oblivious to the existence of justice research (presumably a rather large number) can be assumed to at least occasionally contemplate questions of justice in virtually any area of life. Looking past the subject of this thesis, people can ponder the fairness of societal issues like racial discrimination and gender inequality, personal issues like their salary, taxes, the way they are treated by a host of others ranging from distant authorities to their closest friends, but also things as mundane as getting a parking ticket or being in the slowest line at the supermarket checkout. As a consequence of the ubiquity of justice-related concerns, we can assume that everyone has a more or less concrete notion of what is and what isn't just. Therefore, approaching the subject of justice perceptions from a scientific perspective requires a clear account of what is actually meant by "justice".

2.1 Justice theory: overview

The main theoretical underpinnings of this thesis are derived from organizational justice research. In contrast to the philosophy of justice, which can trace its heritage as far back as Plato's *Republic*, empirical justice research emerged in the second half of the 20th century (Sabbagh and Schmitt 2016). The philosophical discourse on the nature of justice has traditionally been focused on questions of which norms and principles represent a state that can be described as just (Rawls 1971). Conversely, organizational justice research takes an empirical approach by asking how these principles manifest themselves in everyday reality, about their concrete relevance for individuals, organizations, and society (Liebig and Sauer 2016). In particular, this field is concerned with how actors perceive the fairness of the allocation of various resources as well as the procedures that govern these actions (Greenberg 1990). These perceptions have been shown to have far-reaching consequences for individual attitudes and behavior. In workplace settings, justice perceptions have been linked to outcomes such as organizational citizenship behavior

(OCB) (Konovsky and Pugh 1994; Moorman 1991), trust (Colquitt and Rodell 2011; Korsgaard et al. 1995), turnover intentions (Daileyl and Kirk 1992; Poon 2012; Simons and Roberson 2003), and counter-productive work behavior (CWB) (Fox et al. 2001; Skarlicki and Folger 1997). While these examples represent only a small selection of areas in which the importance of justice perceptions has been explored, the main point should be clear: Individuals care deeply about justice, and will show strong reactions if they feel that they are denied justice.

But what does it mean for a thing to be "just"? Organizational justice research was originally grounded in *equity theory* (Adams 1963, 1965; Homans 1961). The main tenet here is that in order to be fair, the allocated amount of a resource should be proportional to the amount of input the receiving party invested in order to be eligible for reception of the resource (Leventhal 1976). This concept has come to be known under the term *distributive justice* (Deutsch 1985). To give a simple example: The wage an employee receives as compensation for their work should reflect their efforts. Or, in context of the present thesis: The more effort a student puts into an assignment, the higher the grade they should get. Jasso (1978, 1980) provided a formalized approach to quantifying the extent to which a distribution deviates from "perfect justice" by computing the logarithm of the ratio between the actual reward an individual received and the reward they would consider to be just. The greater the discrepancy between, for example, the wage an individual received and the wage they feel they deserved, the greater the perceived injustice. Such feelings are not necessarily the result of inequity. Research has shown that, depending on the situation and the actors involved, distributive principles other than equity might be preferred. For example, if the principle of *equality* is to apply, the allocation of resources should bestow all recipients with the same amount without considering their contributions. The principle of *need* demands that the amount assigned to an actor should be tied to the extent to which the actor needs the resource rather than their efforts (Deutsch 1975).

But no matter which justice principle is deemed to be the most appropriate in a given situation, all investigations of distributive justice have one thing in common: Their focus is on the fairness of the *outcomes* of resource distributions. By doing so, they neglect the fairness of the *procedures* that led to the outcomes (Leventhal 1980). In his critique of the state of justice research, Leventhal (1980) states that resource allocations are complex processes guided by various rules and regulations. The actual allocation of the resource is only the final product of these processes. According to Cropanzano and

Folger (1989), "resentment is maximized when people believe that they *would* have obtained better outcomes if the decision maker had used other procedures that *should* have been implemented" (p. 293, italics in original). Therefore, the outcome-focus of distributive justice only gives an incomplete picture of the areas in which justice can be an issue, which gave rise to the study of procedural justice as a distinct justice dimension (Lind and Tyler 1988). Later on, the scope of justice research has been expanded to also include questions of how individuals are treated in the course of a procedure (Bies and Moag 1986). As a result, current theories of organizational justice commonly distinguish four justice dimensions: distributive justice, procedural justice, interpersonal justice, and informational justice (Colquitt 2001).

2.2 The fairness of procedures

The concept of procedural justice was first devised by Thibaut and Walker (1975) in a legal setting. They found that court rulings are seen as more fair if they follow a fair process – regardless of whether or not the final decision was favorable. They relate the fairness of the process to the extent to which individuals are involved in it. People are more likely to view a procedure as fair if they are involved since this gives them partial responsibility for the result. Thibaut and Walker distinguish two types of involvement. *Process control* refers to influencing the procedure by being granted the possibility to express one's own views and feelings toward the process. *Decision control* refers to direct control of the result (Colquitt 2001).

A few years later, Leventhal (1980) introduced a more refined conception of procedural justice. According to Leventhal, a procedure is perceived as just if it is in accordance with a number of rules. The first rule demands *consistency*, meaning that the procedures applied to distribute resources should be constant across both individuals and time. The *bias suppression* rule demands that decisions should be impartial and not guided by self-interest. The *accuracy* rule states that the distribution of a resource needs to be based on all the information that is necessary to make a decision. The rule of *correctability* addresses the necessity of having procedures in place that enable actors to appeal an outcome. The *representativeness* rule demands that a decision has to reflect the views and concerns of all parties affected by the decision. Lastly, the rule of *ethicality* states that procedures need to be in accordance with the general ethical values of a society.

The above criteria show that there are multiple points of failure through which the fairness of a distributive procedure can be compromised. But why is it important that

procedures are fair in the first place? According to the *instrumental model* of procedural justice that goes back to Thibaut and Walker's (1975) work, fair procedures are important because they ensure that outcomes are fair – even though the outcome might not be favorable. The main point here is that as long as the various elements that make up a distributive procedure can be described as fair, there should be no base on which the fairness of the outcome can be contested. A student might not be happy with the grade they received. But if the grading process was fair, they should accept the grade as such since it is the exact grade they deserve. As will be shown later on, this is of course an idealized scenario, since a perfectly fair grading process is hard to implement, and it can be just as hard to convince students that it is, in fact, fair.

It is also important to consider that procedures tend to be relatively stable over time, meaning that they should not change too much between instances in which they are applied. This provides a sense of stability. If a procedure is fair, there is a good chance that the immediate outcome is fair. However, this can also be extrapolated to indicate that future outcomes that result from the same procedure will likely be fair as well (Tyler 2000; Tyler and Lind 1992). Indeed, Lind and Van den Bos (2002) point out that justice judgments are used to deal with uncertainty. While future outcomes cannot be predicted with absolute certainty, fair procedures enable individuals to better gauge the possible outcomes. In the present context, this means that while students cannot fully anticipate how they will perform in future assignments, the use of fair grading procedures at least assures them that their efforts will not be invalidated by questionable grading practices. Thus, uncertainty can be reduced.

A different perspective is provided by the *relational model*. According to Lind and Tyler (1988), individuals value fair procedures because they convey information about their standing in a group. This approach bases procedural justice on *trust*, *standing*, and *neutrality*. The idea that procedural justice conveys information about how an individual is appreciated makes matters of procedural justice also an emotional issue. However, there are several problems with regard to the conception of the relational model. As Colquitt (2001) points out, proponents of the relational model conceptualize and operationalize procedural justice in a way that subsumes concepts that are better treated separately. *Trust* has emerged as the subject of its own field of research (Dirks and Ferrin 2001; Rousseau et al. 1998; Schoorman et al. 2007), based on which Colquitt describes trust as a correlate of procedural justice rather than a justice criterion. *Standing* appears to be more closely related to the concept of interactional justice as introduced by Bies and

Moag (1986) since it does not address characteristics of a procedure but rather whether individuals are treated with respect in the course of the procedure. Therefore, while this thesis does not try to refute the relevance of procedural justice as a point of reference through which individuals can gather information on their status, the general conceptualization and operationalization will follow the instrumental model.

Some aspects of interactional justice will be considered as well, but they will be treated as a distinct justice dimension. Bies and Moag (1986) relate interactional justice to the fulfillment of the criteria *truthfulness*, *respect*, *propriety of questions*, and *justification*. However, Greenberg (1993) argued that these criteria actually target two rather different things. While respect and propriety of questions refer to the quality of interpersonal treatment during the procedure, truthfulness and justification refer to how individuals are informed about the details of a procedure. Because of that, Greenberg decomposes interactional justice into two separate dimensions: interpersonal justice and informational justice. As we will see later on, assessment-related explanations provided to the students are vital to their understanding of the assessment process, which is why this thesis also focuses on matters of informational justice.

2.2.1 Research on procedural justice in education

This thesis applies the concept of procedural justice to the investigation of the fairness of grading procedures in a higher education context. Since performance assessment and with it the distribution of grades play a major role in the learning experience, it can be safely assumed that students care about the fairness of assessment procedures. But even though the argument that grading procedures should be fair is not likely to encounter resistance, procedural justice is rarely investigated in educational contexts (for a recent overview, see Sabbagh and Resh 2016). What research is available on the subject is plagued by a number of issues that prevent applying the approaches found in these studies in the context of this thesis. For the most part, these issues revolve around a lack of specificity with regard to the procedures whose fairness is of interest. This is compounded by a lack of consistency with regard to how justice perceptions are operationalized.

Some of the earliest investigations of procedural justice in education were made by Chory-Assad (2002) and Chory-Assad and Paulsel (2004). The focus of these studies was on the relationship between procedural justice and student motivation, affective learning, and aggression toward the instructor (Chory-Assad 2002) and between procedural justice and aggression toward the instructor and refusal to respect instructor requests

(Chory-Assad and Paulsel 2004). In order to measure justice perceptions, they used scales consisting of 14 respectively 17 items that address a subject whose fairness the students have to rate. For example: "the course attendance policies", "the grading scale of the course", "the scheduling of exams", "the types of questions on exams", and "the way the instructor calls on students in class". It is a rather strong assumption that all of these procedures are of equal importance to the students. Despite that, the authors combine ratings of these diverse and often unrelated procedures into a single measure of procedural justice. This prevents the identification of procedures where injustice is particularly problematic. Another problem arises from the way justice perceptions are measured. The authors use *direct* measures where students are asked to rate the fairness of the subject of interest. The problem is that direct questioning provides no point of reference as to what the criteria are on which these sentiments are based (Colquitt and Rodell 2015). If a student reports that they hold the assessment process to be unfair, is this because the instructor is biased, exam questions are too hard, or some whole other thing?

Procedural justice was also investigated by Resh and Sabbagh (2014a; 2014b). They investigated effects of procedural justice on liberal democratic orientation and social and institutional trust (Resh and Sabbagh 2014a) and effects of procedural justice on the sense of belonging to school and social and institutional trust (Resh and Sabbagh 2014b). Unlike the studies by Chory-Assad (2002) and Chory-Assad and Paulsel (2004), Resh and Sabbagh (2014a, 2014b) use *indirect* measures of justice perceptions. This means that participants are not directly asked about the fairness of a procedure, but rather about the different criteria that characterize a fair procedure (Colquitt 2001). For example, the item "School rules and regulations apply equally to all students in our school" targets the consistency with which rules and regulations are put to use. While this mitigates some of the issues related to direct measures, an important problem remains: It is not particularly clear which "rules and regulations" this refers to, which again leads to a lack of specificity.

A similar issue is apparent in a recent study by Kazemi (2016). He investigated effects of procedural justice on motivation to study and achievement. The theoretical foundation of his measure of procedural justice is given explicitly. It addresses Leventhal's (1980) rules of bias suppression and consistency. However, his six-item instrument contains three items that assess bias suppression with regard to grading procedures, while the three items targeting consistency refer to classroom policies, particularly to how

students are treated for being late. Therefore, the conceptual clarity of this scale is undermined by the fact that both justice criteria refer to different procedures – what if grading is free of bias, but inconsistent? This possibility cannot be captured by applying different criteria to different procedures and treating the result as a single construct. Further, even though it is possible to use such a measure for the prediction of an outcome, the interpretation of the results is impeded by the theoretical ambiguity: Are unfair grades and unfair lateness policies equal contributors to the observed effects?

These examples point to an important issue: Not only is there a lack of research on procedural justice in education, what little research is available also lacks consistency regarding how procedural justice is conceptualized and operationalized. Thus, before we can move on to discuss the research agenda of this thesis, it is necessary to provide a focused and comprehensive account of the theoretical underpinnings on which it rests. Let us first consider the conceptualization of procedural justice. The first question should be what the procedures are whose fairness is being investigated. In the present case, the subject of interest are the procedures used in grading student performance. As the examples from previous research show, there are many other procedures that can potentially affect the learning experience such as course attendance policies or the scheduling of exams. The exclusive focus on grading procedures is not meant to insinuate that other procedures are irrelevant. Rather, it is argued that focusing on a single, well-defined procedure enables us to define the mechanisms that affect both the outcomes and the formation of justice perceptions.

This is not possible if different procedures are blended into a single measure of a general sense of procedural justice. It can be argued that any correlation that can be established between a specific outcome and an unspecific procedure might be sound from a statistical point of view, but lacks with regard to the theoretical implications. The crucial point here is that any attempt to influence – i.e. improve – justice perceptions will be cut short if there is a lack of clarity regarding which procedures are the source of the sense of injustice. Since justice research is usually focused on outcomes rather than antecedents of justice perceptions, this issue is generally overlooked. However, this thesis aims to make a strong case for abandoning the preoccupation with outcomes in favor of an integrated approach that also asks why students question the fairness of the assessment process and how these judgments are formed. Therefore, procedural justice is conceptualized as the fairness of a specific procedure as opposed to an overall sentiment toward an amalgamation of different procedures.

With regard to the operationalization, the above examples from previous research show that there is no established way to measure perceptions of procedural justice in education. Even though a wide range of instruments has been developed and thoroughly tested over the years in other contexts, these are not presently used in educational research. However, it can be argued that the criteria used to define the fairness of procedures such as pay raises should also apply to the grading process. Thus, it should be possible to translate existing instruments to the subject of this thesis. Out of the available instruments, the one proposed by Colquitt (2001) is the most prevalent and also the one that most closely corresponds to the aims of this thesis. One of the greatest advantages of this instrument is that each item can be unambiguously assigned to one of the theoretical concepts that define the fairness of a procedure. These criteria are introduced in the following paragraphs.

2.2.2 Characteristics of a fair grading process

What actually defines a fair procedure? In this thesis, the fairness of the grading process is based on the fulfillment of six criteria that reflect the works of Thibaut and Walker (1975) and Leventhal (1980). The criteria are Thibaut and Walker's (1975) process control and decision control and Leventhal's (1980) rules of correctability, bias suppression, consistency, and accuracy. Note that this thesis does not claim to employ a fully exhaustive list of criteria that could relate to the fairness of procedures. It is entirely possible that fairness perceptions are also influenced by criteria not included in this selection. Nevertheless, it is argued that the chosen criteria cover a vast spectrum of fairness-related issues and are thus assumed to provide a comprehensive picture that is deeply grounded in justice theory. As will be seen, these criteria are easily translated to the grading process as they address a number of issues that are common themes in the literature on assessment of student performance. However, while the higher education literature acknowledges the importance of these criteria, it does not usually connect them to the general field of organizational justice research. The following paragraphs discuss these criteria and their implications for the fairness of grading procedures in greater detail.

The first two criteria of interest are process control and decision control. *Process control* ensures that the affected parties are allowed to present their own views and arguments regarding a procedure. By doing so, they are given some measure of indirect control over the outcome. Conversely, *decision control* does not refer to influencing a procedure, but rather the outcome itself (Colquitt 2001). In the present context, process control

means that students have a voice in determining the grounds on which their work is to be judged. For example, a student can feel that they have contributed a lot in class during the semester, and that this should be reflected in their grade. Decision control means that students have a voice in determining the actual grade.

At first glance, this might appear more like a nuisance than something that is desirable in a higher education context: Should students be involved in deciding the grading criteria? Should they really approach their instructors and demand a certain grade? Admittedly, this could cause some problems, especially if the students confront their instructors with demands that are essentially unfounded. Despite that, it can be argued that granting students process control and decision control can alleviate feelings of injustice. The more involved students are in the grading process, the more responsibility they take for both the process and the result. Due to this increased responsibility, it is assumed that they are more likely to perceive the whole experience as fair. But if the students' contribution is limited to handing in the assignment, all responsibility ultimately lies with the instructor. Students can be under the impression that the grade does not really represent their effort because the instructor did not give them the chance to defend their position. As a consequence, students might feel that grades are forced onto them.

Greater involvement can also equip students with knowledge necessary to self-regulate their learning efforts. Through engaging in dialogue with instructors about grading criteria and grading decisions, students develop a deeper understanding of what constitutes good work. Even if an instructor is not convinced by a student's arguments, the student can use this knowledge in future assignments: Next time, try something different. This is reminiscent of *formative assessment*, a concept to which a large segment of the literature on assessment in higher education is dedicated (López-Pastor and Sicilia-Camacho 2015; Pereira et al. 2015). With regard to student involvement in the assessment process, Deeley and Bovill (2015) argue that a staff-student partnership promotes assessment literacy, a basic precondition for students to improve in future assignments.

Apart from having control over process and decisions, the fairness of a procedure also depends on the existence of some means of retroactively challenging a decision. This is addressed by Leventhal's (1980) rule of *correctability*. This criterion demands that individuals are to be given the right to appeal should they have the feeling that errors have been made. In the context of this thesis, this means that students should have the option to dispute a grading decision if they have reason to believe that the decision was flawed. Attempts to dispute a grade can be motivated by different reasons. For one, it is possible

that a grade resulted from technical errors. A good example for this is an instructor making mistakes when adding up points in an exam. Such errors are easily verified and corrected. But beyond that, it is possible that students have more fundamental concerns regarding an instructor's handling of the grading process. This can happen if the instructor infracted one of the justice criteria to be discussed next: bias suppression, consistency, and accuracy.

The rule of *bias suppression* states that the decision-making process must not be guided by the personal self-interest of the decision-making entity (Leventhal 1980). This ensures that decisions are not corrupted by favoritism or, arguably worse, a dislike for certain individuals, be it on a personal level or on grounds of gender, race, or any other kind of ascribed characteristic. But regardless of the reasons for biased treatment or whether it affects a particular individual in a positive or negative way, the notion that any such sentiment should not influence a procedure will likely find little objection. Grading procedures are certainly no exception in this regard. If grades are to serve as indicators of student effort and knowledge, the decision-making process has to be impartial. The more biased an instructor, the more grades become representations of their sentiments toward their students as opposed to measures of academic achievement. Therefore, a lack of objectivity compromises the validity of the assessment process. It is obvious that this should be seen as unfair by the students, since receiving a good grade becomes dependent on factors that are outside their sphere of influence.

Next, the *consistency* rule demands that resources should be distributed according to uniform standards (Leventhal 1980). In the present context, this means that the standards used to judge the quality of student work should not show pronounced differences. This thesis argues that consistent grading procedures are seen as fair because they make the assessment process reliable. There are two points in which grading procedures can deviate from this ideal. First, there can be differences *between* instructors. Some instructors can be rather lenient, i.e. giving good grades as a matter of habit. Others can be very strict, demanding great efforts from their students and holding even minor mistakes against them. As a consequence, the exact same work can be judged favorably by one instructor while another considers it to be largely disappointing (Wolfe 2004). This can also happen if different instructors apply different criteria (Hand and Clewes 2000). As an example, some instructors might give similar weight to writing style as they give to content and structure when judging the quality of an essay while others might not care about style at all. Again, the same efforts could be awarded disparate grades.

Second, standards can vary *within* instructors. This is the case when instructors change their standards over time. For example, an instructor who was previously known to give primarily good grades can start to favor a stricter approach, awarding poorer grades than the students would have expected. A more severe variant of this would be an instructor who does not have any particular standards in the first place, meaning that their approach to grading varies from assignment to assignment and from student to student. But regardless of whether inconsistencies occur between or within instructors, it seems reasonable to expect that students should feel injustice if their work is held to shifting standards. Of course, given the open nature of some assessment formats (e.g. essays), achieving perfect consistency seems hardly possible (Bloxham et al. 2011; Bloxham et al. 2015). Despite that, there have been attempts to devise frameworks for training instructors in the application of consistent grading strategies (Chen et al 2016).

Finally, this thesis holds the fairness of a procedure to the extent to which the decision-making process is based on *accurate* information. According to Leventhal (1980), a procedure lacks fairness if it relies on incomplete or inaccurate information, or if the person in charge of the procedure is not qualified for the task. Applied to grading procedures, this means that instructors have to take special care to ensure that assessment accurately reflects student knowledge of the subject matter. There are several ways in which this rule can be violated. Leventhal (1980) gives exams that are too hard as an example. In a similar fashion, questions regarding a topic that was never discussed in class are unlikely to accurately capture student knowledge since students had no chance to prepare for this topic. It is also possible that an instructor misses an important point a student made in an essay and gives a lower grade because of that, just as it is possible that an instructor finds a correct answer to be wrong because their own knowledge of the subject matter is inaccurate. It is obvious that students should feel injustice if grading is based on distorted information instead of their actual knowledge. Regardless of whether this happens due to a lack of experience or if it merely points to negligence: If the assessment process fails to capture student knowledge of the subject, it fails to measure what it intends to measure. Therefore, the resulting grade will not be substantiated, rendering it invalid.

2.2.3   Control and validity - two facets of procedural justice

Looking closely at the above criteria, it becomes clear that even though all of them address various properties of the same procedure, each property can be assigned to one of

two distinct categories. The first three criteria – process control, decision control, and correctability – address student involvement in the grading process. The other three criteria – bias suppression, consistency, and accuracy – aim at the degree to which grading procedures are capable of producing results that can be considered valid.

The sentiment that grading procedures are unfair can thus be caused by both a lack of control and by validity-related concerns. In the organizational justice literature, this distinction is usually not made. Nevertheless, this thesis argues that separating control-related from validity-related aspects of procedural justice has distinct advantages. It is possible that grading procedures are both valid and give students a decent amount of control. However, it is equally possible to encounter situations in which students are largely kept from being involved in the grading process while at the same time, the procedures are highly objective and adhere to strict standards. Conversely, students could be closely involved in the various steps of the grading process while this involvement is inconsistent and shows signs of favoritism. Since control does not necessitate validity and validity does not necessitate control, combining both aspects of procedural justice in a single construct results in a loss of information. Hence, this thesis distinguishes the two facets of procedural justice on both the theoretical and the empirical level. The terms control-related procedural justice (PJ-C) and validity-related procedural justice (PJ-V) will be used as designations.

2.3   Knowledge or conjecture? The role of information in shaping justice perceptions

At this point, it is required to emphasize the fact that in the present context, justice perceptions are viewed from the *students' perspective*, and thus based on their own interpretation of the situation. Such interpretations are inherently subjective so that there is a distinct chance that they do not accurately reflect reality. From the student perspective, grading criteria are often unclear (Carless 2006). This lack of transparency can impede the perceived legitimacy of the assessment process (Lizzio and Wilson 2008). To give an example, a student might feel that their instructor holds a grudge against them, giving them worse grades than they deserve. But can one infer from this that the grading process is actually biased? It is entirely possible that this is merely a suspicion on the student side that is not grounded on actual teacher misconduct, but rather on pure conjecture. With regard to the consequences of perceived injustice, this point is rather moot: As long as grading appears unfair, students are prone to act accordingly. But what can be done to

avoid students getting the wrong impression in the first place? This issue is covered by the concept of *informational justice*.

Informational justice is based on Greenberg's (1993) work. In the present context, informational justice (IJ) relates to explanations of the grading process provided to students by their instructors. Altogether, the goal should be to provide feedback that makes the assessment process transparent. This is the basic condition that needs to hold if students are to accurately assess the fairness of the grading process. In absence of sufficient information, some sort of "shooting the messenger" scenario could occur: Students could try to legitimize a sub-par grade by blaming the result on the instructor being biased, incompetent, etc. instead of connecting it to their own lack of effort.

Similar to the two facets of procedural justice, this thesis uses a number of criteria to indicate the fairness of information regarding grading procedures: extent, comprehensibility, and timeliness of feedback information. These are derived from Colquitt's (2001) organizational justice scale. In turn, Colquitt based his informational justice items on the works of Bies and Moag (1986) and Shapiro et al. (1994).

The first criterion is the *extent* of information a student receives. The relevance of this category is rather obvious: Feedback that merely scratches the surface is of little practical use for the students. If there is legitimate need to receive feedback, short explanations will not be enough. At worst, short, generic comments offer no perceivable value over not giving any feedback at all (Ferguson 2011). The available information is the primary reference on which all justice judgments are based. The more room feedback leaves for interpretation, the greater the risk that student perceptions of the assessment process deviate from reality.

Next, feedback should be *comprehensible* if it is to be seen as fair. Students show problems both in understanding what feedback says on the semantic level and in understanding what they are supposed to do based on the feedback (Winstone et al. 2016). If feedback is hard to understand – whether because it is too vague and/or makes excessive use of academic jargon – chances are lower that students can benefit from it (Higgins et al. 2001). At worst, feedback that is hard to understand can achieve the exact opposite of what it is intends to do. If students suspect that the explanations actually obscure more than they clarify, feelings of injustice might even be aggravated.

Finally, the *timeliness* with which students receive explanations is considered an important property (Bayerlein 2014; Gibbs and Simpson 2004). Students generally appear to prefer to receive feedback as soon as possible (Mulliner and Tucker 2015). The less

time between assignment and feedback, the greater the salience of the subject in the students' cognition. But the more time passes, the more an assignment fades into irrelevance in order to make room for more current, pressing issues (Denton et al. 2008). This is particularly problematic in terms of the relationship between IJ and PJ-V. While feedback can be used to dissipate validity-related concerns, it can be argued that there is a decent chance that students will stick to their incorrect assumptions if the clarifications come too late. And even if an instructor ultimately manages to convince a student that their concerns were unfounded, the waiting period will still be accompanied by feelings of injustice.

Matters of informational justice have found even less attention in educational studies than procedural justice. Thus far, only one study has explicitly targeted informational justice among students, and this study is focused on secondary education rather than higher education (Kazemi 2016). Nevertheless, research on *assessment feedback* is actually one of the most prominent areas of interest in the higher education literature (for a recent review, see Evans 2013). Research in this field has identified assessment feedback as one of the most essential factors that influence student learning. Quality feedback helps students to identify their strengths and weaknesses, which provides a point of reference for improvement in future assignments (Black and Wiliam 1998; Hattie and Timperley 2007; Nicol and Macfarlane-Dick 2006). This is conceptualized as an iterative procedure in which the feedback for each assignment provides input for the next, which promotes a continuous refinement of skills through which assessment gains a formative function (Evans 2013).

The majority of sources cited in the above discussion of extent, comprehensibility, and timeliness of feedback actually belongs to this field, which shows that the importance of each informational justice criterion is generally recognized in the feedback literature. But despite these similarities, the feedback literature is primarily centered around questions of feedback *effectiveness*, and thus remains disconnected from justice research. Nevertheless, since the idea that feedback can serve to legitimize grading decisions is generally accepted (Ferguson 2011; Lizzio and Wilson 2008), a straightforward link to the concept of informational justice can be established.

3 Research agenda

The previous chapters established the theoretical framework of this thesis. The individual criteria that characterize fair procedures were introduced along with a discussion of their role in the grading process. This allows us to specify the mechanisms through which justice perceptions influence student attitudes toward their studies. It further allows us to target the formation of the justice perceptions themselves. The theoretical framework is now applied to define a research agenda for the empirical investigation of justice perceptions in a real world higher education setting. Altogether, three studies were conducted that target the different elements of this agenda. The following chapters introduce these studies in greater detail.

3.1 Consequences of unfair grading procedures

3.1.1 Research question

The aim of Study 1 is to demonstrate how student attitudes toward their studies are affected by the fairness of grading procedures. Specifically, this study explores the relationship between perceptions of procedural justice and *dropout intentions*. The choice to investigate dropout intentions as an outcome of procedural justice perceptions was motivated by two reasons. First, regardless of the continuously high number of students who prematurely terminate their studies every year, research on dropout from higher education in Germany remains surprisingly sparse (Heublein 2014). This is particularly true for studies that go beyond the purely descriptive level by using multivariate analyses (Blüthmann et al. 2008; Georg 2009). Second, studies in workplace settings have provided evidence for a connection between justice perceptions and turnover intentions (Daileyl and Kirk 1992; Poon 2012; Simons and Roberson 2003). Since the act of quitting one's job can be considered as functionally equivalent to university dropout, it is plausible to expect a relationship between perceived fairness and dropout intentions as well.

While previous research on procedural justice in education has investigated outcomes such as sense of belonging that could potentially be related to dropout (Resh and Sabbagh 2014b), dropout intentions themselves have not yet been investigated as an outcome. Therefore, it is necessary to first establish the theoretical mechanisms through

which the perceived fairness of grading procedures is expected to affect dropout intentions.

3.1.2 Theoretical background

The majority of research on dropout from higher education in English-speaking countries relies heavily on Tinto's (1975) seminal model of student dropout. Tinto conceptualizes dropout as the result of a failure to develop a sufficiently high level of integration into the social and academic system of the university. A lack of integration causes the cost-benefit ratio of continuing one's studies to fall below that of a possible alternative, making dropout the most viable pathway. But despite the ubiquity of Tinto's model, its applicability to the German context is somewhat questionable. For example, his idea of social integration assumes an institutional framework that is not present in most German universities (e.g. on-campus housing). But more importantly, his theory does not account for the perceived *probability of success*, one of the core elements in sociological models of educational decision-making (Breen and Goldthorpe 1997; Esser 1999).

A nationwide study on dropout in Germany by Heublein et al. (2010) has identified student achievement as the most important factor in deciding whether or not students prematurely leave their studies, followed by financial problems, a lack of motivation, and poor study conditions. The immediate importance of achievement for the probability of success is obvious. If a student receives continuously poor grades, the feasibility of further investments in the current studies decreases until a point is reached where the benefits of dropping out outweigh those of continuing.

Study 1 adds student perceptions of the fairness of grading procedures to this equation. The central argument is that fair procedures reduce uncertainty with regard to the probability of success. Grading procedures are the means through which the students' knowledge and efforts are converted to abstract quantities in the form of grades. The more closely grading procedures adhere to the various justice criteria, the greater the chance that students receive a reward that reflects their efforts. If, on the other hand, grading procedures violate the justice criteria, the connection between effort and success is weakened. Whether or not students receive a grade that promises success becomes dependent on factors that are outside the students' sphere of influence. Therefore, how students perceive their achievement and the implications for their chances to succeed is expected to depend on the fairness of grading procedures.

Building on this, Study 1 formulates hypotheses that specify how perceptions of PJ-C and PJ-V are incorporated into the students' deliberations on whether they should quit their studies in favor of an alternative option. In order to add another layer to the investigation, this study also considers potential differences in the effects of fairness perceptions according to the students' social background.

In doing so, this study aims to demonstrate the advantages of employing a focused, well-defined conception of procedural justice. The focus is on the fairness of *grading procedures* because of their role in distributing the very resource that is the most salient indicator of educational success: grades. This allows for a clear specification of the mechanisms through which adherence to respectively violation of the justice criteria influence dropout intentions. It is possible that dropout intentions are affected by other distributive procedures as well. Nevertheless, since these other procedures would govern the distribution of resources *other* than grades, it is crucial that they are treated as distinct constructs if they are to be investigated – both theoretically and empirically. Combining various procedures to a global sentiment of procedural justice would preclude the identification of the contribution of each procedure to the outcome of interest. In a way, the present research goes even further by not only focusing on a specific procedure, but also distinguishing the fairness of two separate facets of the same procedure: control and validity.

3.1.3   Methodological approach

The empirical implementation of this research agenda necessitated the development of an instrument for measuring students' justice perceptions. The discussion of the shortcomings of previous research on procedural justice in education shows that the instruments used in these studies are not suited for the present purpose. A wider selection of instruments is available in the organizational justice literature, out of which the choice was ultimately in favor of the scale developed by Colquitt (2001). The reasons for this decision have been mentioned above: 1.) High conceptual clarity of the justice criteria. Each item targeting procedural justice can be directly linked to one of Thibaut and Walker's (1975) and Leventhal's (1980) justice criteria. 2) Unambiguous assignment of criteria to justice dimensions. Measures of procedural justice that are based on Lind and Tyler's (1988) group value model usually also include items that more closely reflect interpersonal treatment instead of characteristics of procedures. 3) Widespread adoption in organizational justice research. Even though the recognition of the scale does not extend to

educational studies, Colquitt (2001) himself tested his instrument on a sample of university students, so that suitability for the present research context can be assumed.

Colquitt's items were translated to German. The translation was partially based on Maier et al. (2007). The wording was adapted to specifically target instructor behavior when grading assignments. The original German item formulations are presented in Table 4-1; English translations are given in Table 5-3. The main data collection was done as part of the *CampusPanel*, a large-scale online survey that was conducted in late 2013 at the University of Tübingen, Germany. The CampusPanel is an interdisciplinary effort with contributions from sociologists, social psychologists and educational scientists. A detailed account of the development and implementation of the CampusPanel is provided in the official data documentation (Lang and Hillmert 2014).

3.2 Justice perceptions and the academic environment

3.2.1 Research question

Study 2 marks the transition from using justice perceptions as predictors of an outcome to treating them as the outcome itself: What determines whether or not grading procedures appear fair from the students' point of view? This shift in perspective is rarely found in previous research. The consequences of justice perceptions are usually the primary subject of interest, to the detriment of asking *why* something is seen as fair (or unfair).

This thesis argues that the incessant focus on outcomes of justice perceptions limits the usefulness of the knowledge derived from such research. This can be illustrated by the example of grading procedures and dropout intentions. Under the assumption that retention rates should ideally be as high as possible, the finding that students base their plans to end their studies on the fairness of grading procedures should favor the conclusion that grading procedures need to be fair. Nevertheless, this type of conclusion is seldom seen as an incentive for asking about the reasons behind perceived (in-)justice. Therefore, research that is purely outcome-focused is limited in that it describes problems without attempting to provide solutions.

In their recent review on the state of research on justice in education, Sabbagh and Resh (2016) explicitly point out the need to investigate the role of structural characteristics of educational organizations in influencing justice perceptions. Study 2 takes on this task by focusing on the role of the *academic environment*. The basic idea here is that the approaches to teaching, methods of assessing student performance, and the organization

of curricula vary heavily between departments, and with them the experiences students make (Neumann et al. 2002). It is now argued that these differences are also reflected in the students' justice perceptions. The aim of Study 2 is thus to establish a link between individual-level justice perceptions and structural characteristics on the department level.

3.2.2   Theoretical background

The idea of exploring justice perceptions from an institutional angle is derived from the *justice climate* approach. For a long time, justice perceptions were viewed as individual level phenomena. This assumption was challenged by the justice climate approach, which emphasizes that in order to arrive at a thorough understanding of justice perceptions, the embeddedness in their respective social and institutional contexts needs to be accounted for (for a recent review, see Li et al. 2015). The basic idea here is that members of the same organizational unit are subject to the same rules and regulations and are therefore bound to make similar experiences (Mossholder et al. 1998; Naumann and Bennett 2000). Since individuals in a group interact with one another, it is assumed that they also exchange their views regarding justice-related matters. Through these exchanges, justice perceptions converge over time, giving rise to a group-specific justice climate (Degoey 2000).

Applied to the subject matter of this thesis, it is expected that justice perceptions should vary between departments. The reasoning here is pretty straightforward: Student experiences of the assessment process within a department are bound to be more similar to each other than to experiences made by students in other departments. But while justice climate research generally acknowledges the importance of being exposed to the same contextual conditions, there have been no attempts to explicitly model these conditions in order to explain differences between groups. Study 2 aims to take on this issue. The argument is that students' justice-related experiences are affected by the method used to assess student performance as well as the format in which instruction takes place.

As for the assessment method, this study distinguishes between *essays* and *exams*. Both assessment formats provide for very different experiences. In the present context, the crucial distinction is the degree to which assessment and grading are standardized. Exams such as multiple-choice questionnaires are generally less open than essays. The quality of student performance in an exam is assessed along a narrow, clearly defined set of criteria such as points awarded for right answers. This is hardly possible with essays, where judging the quality is a complex process that leaves more room for interpretation

(Bloxham et al. 2011; Norton 1990; Orr 2007). Thus, both assessment formats suggest decidedly different approaches to grading, which is expected to have consequences for the students' justice perceptions.

Whether instruction takes place in *seminars* or in *lectures* informs us about how students interact with their instructors. Students in traditional lectures usually remain in a passive role. On the other hand, seminars promote direct interactions between students and faculty (Cuseo 2007; Severiens and Schmitt 2009; Severiens et al. 2015). It is argued that the mode of interaction suggested by the instruction method affects the chances that the students' justice-related expectations are met.

Neither of those methods is expected to fully *determine* justice perceptions. However, it is assumed that the choice of a certain assessment and instruction method gives a specific *direction* to the experiences students are likely to make. Each department is located somewhere along a continuum that ranges from purely essay-based to purely exam-based assessment. The exact location of a department between those extremes determines the extent to which student experiences of the assessment process are defined by the properties of either essays or exams. The same is true for the instruction method, where departments fall somewhere in-between purely seminar-based and purely lecture-based.

Study 2 uses the proportion of essays relative to exams and that of seminars relative to lectures as indicators for how assessment and instruction appear from the student perspective. Based on this conception, a number of hypotheses are developed that relate student perceptions of PJ-C, PJ-V, and IJ to assessment method and instruction method in a department. The argument is that the diverging properties of essays and exams (respectively seminars and lectures) have specific implications for how the assessment process conforms to the underlying justice criteria from the students' perspective. This leads to the emergence of department-specific justice climates. Part of this study also explores the possibility that the magnitude of the effects of the institutional environment could depend on the students' social background.

3.2.3 Methodological approach

The necessity of applying a clear specification of justice criteria and of the procedure of interest is even more apparent in the context of Study 2 than it was in Study 1. It is hardly possible to derive any meaningful hypotheses regarding factors that influence how a procedure is perceived without an explicit definition of *which* procedure is meant. If the goal is to relate institutional structure to the fulfillment of justice criteria, there cannot be any

ambiguity as to what these criteria are. Study 2 complements Study 1 not only by attempting to show how attitudes that ultimately influence dropout intentions are formed, but also by using the CampusPanel as the primary data source. Therefore, consistent justice measures are used between Studies 1 and 2, which contributes to the goal of creating a coherent framework for the investigation of justice perceptions in education.

Individual-level data from the CampusPanel is supplemented with department-level data from the university's online course catalog. Based on this data, multilevel mixed models are used to first quantify the differences in perceptions of PJ-C, PJ-V, and IJ between the 48 departments in the sample. Afterwards, the extent to which these differences can be attributed to the characteristics of assessment and instruction is explored.

3.3 Exploring the justice judgment process

3.3.1 Research question

While Study 2 is centered around the *why* of justice perceptions, Study 3 has its focus on the *how*. Thus far, this thesis has defined a procedure to be fair if it meets a number of different criteria. This provides for a high level of conceptual clarity, but also rests on the assumptions that within each set of criteria assigned to a justice dimension 1) each criterion is an equal contributor to the sense of justice; 2) the contribution of each criterion is constant across situations; 3) the contribution of each criterion is constant across institutional contexts. Since these assumptions appear rather strong, an empirical investigation of how students use various properties of a situation to arrive at a judgment regarding the fairness of that situation is warranted. The goal of Study 3 is to explore this process in a comprehensive, integrated framework.

The focus of Study 3 is on informational justice, which is motivated by two points. First, informational transparency is the basic property that enables students to judge the fairness of the assessment process. As long as grading procedures remain arcane, justice perceptions will be based on supposition. Therefore, an improvement in the students' understanding of assessment and grading can serve as an additional line of support in the fight against perceived injustice alongside measures that target the institutional structure. Second, issues of assessment feedback are the subject of an ever-expanding body of literature in higher education research. A common theme that permeates a large share of these studies is student dissatisfaction with the feedback they receive (Ferguson 2011; Hounsell et al. 2008). The larger the dissatisfaction, the greater the risk that students are

unwilling to accept the feedback and engage with it (Orsmond and Merry 2011; Small and Attree 2015). This is seriously detrimental to the role of feedback in advancing student learning and making assessment transparent (Evans 2013).

The question is thus: What defines fair feedback from the students' perspective? This research question is divided into three parts. First, this study investigates the relative weight students assign to various properties of a feedback situation when judging the fairness of that situation. The second part is concerned with how the demand for different types of feedback information varies between assessment situations. Finally, the third part explores differences in feedback-related expectations between students from different disciplinary backgrounds.

3.3.2   Theoretical background

The above research questions point to issues that are largely ignored in empirical justice research. The process through which justice criteria are weighted against one another when evaluating the fairness of a situation was already discussed by Leventhal (1980). Unfortunately, the impact of his justice judgment model was considerably smaller than that of his definition of the procedural justice rules. This arguably poses some problems with the dominant practice of inferring justice perceptions from indirect, item-based measures. Usually, survey participants are presented a list of items that assess the extent to which a procedure is compliant with the various rules that characterize a fair procedure (Colquitt and Rodell 2015). For example, the indirect measure of informational justice used in the CampusPanel includes three items that ask students if the feedback they receive is extensive, comprehensible, and timely, respectively. The problem here is not that these questions might not be indicative of informational justice – given the theoretical discussion on which they are based, we will assume they are. Rather, it is the assumption that each of the justice criteria addressed by these questions is *equally important* that warrants special attention. In case of informational justice, should we really assume that, for example, the extent of information is equal in rank to feedback timeliness? Questions such as this are usually not accounted for when combining the ratings of individual items to compute a score of justice perceptions.

Another important point to consider is the actual *demand* for feedback. For example, if a student reports to have received short, vague feedback, we would interpret this as a lack of fairness as it violates the justice criteria. But what if this student did not expect

to receive much feedback in the first place? Would this still represent injustice? The argument here is that the demand for assessment feedback is not constant, but reflects the specific needs created by an assessment situation. Indeed, there is some evidence in the literature that suggests that students care less about feedback in case they have gotten a good grade (Crisp 2007). Again, this causes problems with how justice perceptions are commonly assessed. On top of that, the question of what is fair depends on the individual frame of reference. It can be argued that disciplinary differences in curricula and the type of knowledge taught to the students lead to very different experiences with feedback (see Study 2), which could also affect informational demand.

3.3.3   Methodological approach

The most viable solution for approaching these issues is provided by a *factorial survey experiment*. In a factorial survey, participants are presented with short descriptions of hypothetical scenarios, called *vignettes*. Vignettes describe multiple characteristics of the scenario (*dimensions*) whose values (*levels*) systematically vary between vignettes. Each respondent is tasked with rating a number of vignettes based on the levels of its dimensions (Auspurg and Hinz 2015). In the present context, the vignettes depict situations in which a student has received feedback from their instructor following a graded assignment. The vignette dimensions give information on various characteristics of the feedback itself as well as on the situation in which feedback is given. Based on this information, the students rate the fairness of the feedback in the vignette situation. Afterwards, the vignette ratings are used as the dependent variable in regression analyses while the vignette dimensions are used as the independent variables. This enables us to retrace the process through which individual dimensions are concurrently weighted against each other when judging the fairness of the vignette situation.

It is important to point out that this research agenda could not be approached by using item-based measurements in a "traditional" survey. If students were presented a list of individual items whose importance for a fair feedback process they have to rate, there would be no constraints that could keep them from overstating their demands by claiming high importance for *all* items. Since item ratings are independent of each other, it is not possible to make any inferences regarding the relative importance of a particular item. This issue is easily circumvented by using a factorial survey (Liebig et al. 2015). To give an example, a vignette describes a situation in which extensive explanations regarding

the grading criteria were provided. But at the same time, the feedback was hard to understand. This forces the participants to make a trade-off: While extensive feedback is positive and should therefore increase the vignette rating, the poor comprehensibility should subtract from it. In another vignette, feedback might be short, but is easy to understand. Thus, the extent of feedback should subtract from the rating, while comprehensibility should add to it. By comparing the ratings produced by different combinations of vignette dimensions, it is possible to quantify the relative contribution of each dimension to the overall rating of informational justice.

In addition, the inclusion of vignette dimensions that hold contextual information enables us to show how informational needs vary according to factors such as the grade a student received. Again, this would not be possible in the context of a traditional survey (Liebig et al. 2015). One could specify a grade in the introductory text to an item-based measure or even in the formulation of the items themselves. However, all this would do is ensure that the results are valid for the grade specified. This does not grant any insights into the actual role of the grade, because if the grade is not varied, there is no way of knowing whether the results would have been different for another grade.

Of course, controlled laboratory experiments could be conducted to alleviate some of the limitations of traditional surveys since manipulations of factors such as grades are easily implemented. However, this approach is generally plagued by generalizability problems due to small sample sizes and selective recruitment of participants (Liebig et al. 2015). Factorial surveys can be administered via the same channels as traditional surveys (in fact, this is what was done for Study 3). This decreases costs and promises virtually seamless scalability to large sample sizes. In the present case, data is available for 1,129 students from 46 departments of a major German university. Since each respondent rated ten vignettes, results are based on a total of 11,290 vignette ratings. By using a large, heterogeneous sample from many different disciplines, it is also possible to establish the extent to which feedback-related expectations are either specific to a particular disciplinary group or of a more general nature.

4 Study 1: Gerechtigkeit und Studienabbruch. Die Rolle der wahrgenommenen Fairness von Benotungsverfahren bei der Entstehung von Abbruchsintentionen

Burger, R., & Groß, M. (2016). Gerechtigkeit und Studienabbruch. Die Rolle der wahrgenommenen Fairness von Benotungsverfahren bei der Entstehung von Abbruchsintentionen. *Zeitschrift für Erziehungswissenschaft*, *19*(3), 625-647. doi:10.1007/s11618-016-0672-8.

Zusammenfassung   Dieser Artikel untersucht den Einfluss der Fairness von Benotungsverfahren auf die Abbruchsintentionen von Studierenden. Der Fokus liegt auf der Frage, wie Studierende die Notenvergabe aus Sicht der prozeduralen Gerechtigkeit bewerten. Dabei unterscheiden wir zwischen einfluss- und validitätsbezogener prozeduraler Gerechtigkeit. Die Fairness der Benotungsverfahren wird in ein Modell der rationalen Bildungswahl integriert. Wir zeigen, dass Studierende die prozedurale Gerechtigkeit bei der Einschätzung ihrer Erfolgschancen berücksichtigen: Ungerechte Benotungsverfahren erhöhen die Bereitschaft, das Studium abzubrechen. Wir untersuchen auch, inwiefern dieser Zusammenhang herkunftsspezifische Differenzen aufweist. Entsprechende Hypothesen werden anhand von Daten der ersten Welle des CampusPanels geprüft ($n = 1393$). Die Ergebnisse weisen darauf hin, dass Abbruchsintentionen signifikant von der Fairness der Benotungsverfahren beeinflusst werden. Außerdem kann gezeigt werden, dass der Zusammenhang zwischen Gerechtigkeit und Abbruchsintentionen für Studierende mit niedrigem elterlichem sozioökonomischen Status sowie für Studierende mit Migrationshintergrund besonders stark ausgeprägt ist.

Schlüsselwörter   Studienabbruch – Benotungsverfahren – Bildungserwerb – Prozedurale Gerechtigkeit

Abstract   This article investigates the relationship of fair grades and university students' dropout intentions. Our focus is on the extent to which students evaluate the grading procedures as fair from a procedural justice perspective. We distinguish control-related and validity-related aspects of procedural justice. The fairness of grading procedures is integrated into a rational choice model on educational decision-making. We theorize that students consider questions of procedural justice in determining their chances of success. Thus, unfair grading practices lead to an increased risk of the students developing dropout intentions. Further, we expect the magnitude of the effects of procedural justice to be determined by the students' social background characteristics. Data from the first wave of the CampusPanel are used ($n = 1393$). Results show that dropout intentions are significantly affected by procedural justice evaluations. Further, there is evidence that students with low parental socioeconomic status as well as students with an immigrant background are especially prone to exhibit dropout intentions when confronted with low procedural justice.

4.1 Einleitung

Studienabbrüche sind an deutschen Universitäten alles andere als eine Randerscheinung. Von allen Studierenden, die im Jahr 2006 ein Bachelorstudium aufgenommen haben, haben 35% dieses vor Erreichen eines Abschlusses beendet (vgl. Heublein et al. 2012, S. 12). Die Folgen eines Studienabbruches sind vielfältig. Abgesehen von den individuellen Konsequenzen wie etwa höheren Kosten aufgrund einer längeren Ausbildungsdauer dürfen insbesondere die Folgen auf gesamtgesellschaftlicher Ebene nicht übersehen werden. Neben einer ineffizienten Nutzung öffentlicher Gelder zur Finanzierung der Hochschulen führen hohe Abbruchraten auch zu einer Unterversorgung des Arbeitsmarkts mit qualifizierten Fachkräften (vgl. Leszczensky et al. 2009).

Die hohe Prävalenz von Studienabbrüchen steht im Kontrast zu der eher dürftigen Aufmerksamkeit, welche der Thematik in wissenschaftlichen Untersuchungen in Deutschland zuteilwird (für einen Überblick vgl. Sarcletti und Müller 2011). Aus vorhandenen Studien ist bekannt, dass Studienabbrüche als Resultat des Zusammenspiels verschiedener Faktoren verstanden werden können, wobei Leistungsprobleme als häufigster Grund für einen Abbruch angeführt werden (vgl. Heublein et al. 2010). Wir knüpfen an diesem Punkt an, indem wir untersuchen, inwiefern die Intention, das aktuelle Studium vor Erreichen eines Abschlusses zu verlassen, nicht nur durch mangelnde Leistung bedingt ist, sondern auch durch die *Verfahren*, anhand derer die von den Studierenden erbrachten Leistungen bewertet werden.

Die Bewertungsverfahren stehen an der Schnittstelle zwischen Leistung und Erfolg (vgl. Tata 1999) und werden von den Studierenden permanent hinsichtlich ihrer Fairness evaluiert (vgl. Jasso und Resh 2002). Diese Evaluation hängt von einer Reihe von Faktoren ab: Inwiefern sind die Studierenden in den Benotungsprozess involviert? Nach welchen Kriterien wird die Leistung bewertet? Sind die Verfahren objektiv, oder lassen sie Platz für Willkür? Im Folgenden untersuchen wir, wie Bewertungen dieser institutionellen Rahmenbedingungen sich auf das Vorhaben auswirken, ein Studium abzubrechen.

Wir knüpfen dabei an das in der Organisationspsychologie weit verbreitete Konzept der *prozeduralen Gerechtigkeit* an, das einen theoretischen Rahmen zur Beschreibung der Fairness von Verfahren zur Ressourcenverteilung bietet (vgl. Colquitt et al. 2013; Leventhal 1980; Lind und Tyler 1988; Thibaut und Walker 1975). In Arbeitsplatz-

studien haben Gerechtigkeitsevaluationen ein großes Potential für die Erklärung verschiedener Einstellungs- und Verhaltensaspekte gezeigt. Prozedurale Gerechtigkeit stellte sich dabei als wichtige Determinante verschiedener organisationsrelevanter Verhaltensweisen heraus, wie etwa organisationales Commitment (vgl. Konovsky und Cropanzano 1991; McFarlin und Sweeney 1992; Tremblay et al. 2010) oder das Kündigungsverhalten von Arbeitnehmern (vgl. Cohen-Charash und Spector 2001; Daileyl und Kirk 1992; Konovsky und Cropanzano 1991). Gerade der letzte Punkt macht prozedurale Gerechtigkeit bei der Untersuchung von Studienabbrüchen zu einem interessanten Konzept. Unfaire Verfahren schaffen Unsicherheit, was in der Folge die Abbruchsintentionen der Studierenden beeinflussen kann, wobei wir davon ausgehen, dass die Auswirkung prozeduraler Gerechtigkeit auf Abbruchsintentionen mit dem sozioökonomischen Status und dem Migrationshintergrund der Studierenden variiert.

Dieser Artikel erweitert den Forschungsstand in zwei Bereichen. Erstens spezifizieren wir einen theoretischen Mechanismus, über den Gerechtigkeitsevaluationen in Bildungsentscheidungen einfließen. Damit betonen wir die bislang kaum wahrgenommene Relevanz der Fairness von Verteilungsverfahren in Bildungskontexten. Zweitens gehen wir der Frage nach, inwiefern die Reaktionen auf wahrgenommene Ungerechtigkeit sich in ihrem Ausmaß zwischen Studierenden mit unterschiedlichem soziodemographischem Hintergrund unterscheiden. Hier besteht noch erheblicher Forschungsbedarf, weil die situativen Gegebenheiten in diesem Bildungsabschnitt sich stark von Bildungsentscheidungen am Zugang zur Hochschule unterscheiden, welche bislang im Fokus der Forschung stehen (vgl. Müller und Pollak 2010). Insbesondere multivariate Analysen zu den Abbruchsrisiken von Studierenden mit Migrationshintergrund in Deutschland sind bislang nicht vorhanden, so dass über deren Abbruchsmotivationen nur wenig bekannt ist (vgl. Kristen 2014).

4.2   Theorie

Das am weitesten verbreitete Modell zur Erklärung von Studienabbrüchen wurde von Vincent Tinto entwickelt (1975, 1993). Tinto betrachtet die vorzeitige Beendigung eines Studiums als längsschnittlichen Prozess, an dessen Ende eine unzureichende akademische und soziale Integration in die Hochschule die Studierenden zum Abbruch bewegt. Akademische Integration ergibt sich aus den von den Studierenden erbrachten Leistungen sowie deren Lernfortschritten und persönlicher Entwicklung. Soziale Integration folgt aus der Einbindung in Beziehungsnetzwerke mit anderen Studierenden. Insofern Studierende

Probleme bei der Integration in das akademische und soziale System der Hochschule haben, sinkt ihre Verpflichtung dem Ziel gegenüber, das Studium abzuschließen. Sobald die wahrgenommene Kosten-Nutzen Relation des aktuellen Studiums in Folge mangelnder Integration unter der einer alternativen Option liegt, entscheiden sich Studierende für einen Abbruch (vgl. Tinto 1975).

Dem eigentlichen Abbruch ist laut Bean (1982) die *Intention* zum Abbruch vorangestellt. Je mehr eine Alternative zum aktuellen Studium im Studienverlauf an Attraktivität gewinnt, desto stärker die Abbruchsintention und desto höher das Risiko, dass ein tatsächlicher Abbruch erfolgt. Ein tatsächlicher Abbruch ist damit kein plötzliches Ereignis, sondern steht am Ende einer Abfolge negativer Erfahrungen, welche jede für sich die Präferenz für das aktuelle Studium relativ zu anderen Alternativen reduzieren. Im Sinne Tintos kann es sich dabei auch um Probleme bei der Integration in das akademische und soziale System handeln, wobei unser Fokus auf der Bewertung von Studienleistungen als Teilaspekt akademischer Integration liegt.

Eine einzelne schlechte Prüfungsleistung sollte Abbruchsintentionen nur marginal erhöhen und kann durch andere, positivere Erfahrungen ausgeglichen werden. Bei anhaltenden Leistungsproblemen hingegen können die Intentionen einen Punkt erreichen, an dem eine Fortführung des Studiums nicht mehr sinnvoll erscheint. Wir nehmen an, dass die Verbindung zwischen akademischer Leistung und Abbruchsintentionen über die subjektive Einschätzung der Wahrscheinlichkeit vermittelt wird, das Studium mit Erfolg abschließen zu können. Die Bedeutung der Einschätzung der subjektiven Erfolgswahrscheinlichkeit einer Bildungsoption für Bildungsentscheidungen wird auch von der soziologischen Bildungsforschung betont (vgl. Breen und Goldthorpe 1997; Breen et al. 2014; Esser 1999). Sie geht als zentraler Faktor in das Kosten-Nutzen Kalkül ein, das die Wahl spezifischer Bildungsoptionen leitet. Dabei hat das Motiv, in der Generationenfolge mindestens den sozialen Status der Eltern zu erreichen, herkunftsspezifische Bildungsentscheidungen zur Folge (vgl. Breen und Goldthorpe 1997).

In Anlehnung an Tintos längsschnittliche Konzeption von Studienabbrüchen gehen wir davon aus, dass eine kontinuierliche Anpassung der subjektiven Erfolgswahrscheinlichkeiten an die aktuelle Performanz stattfindet. Während des Studiums geben die in studienbegleitenden Prüfungen erreichten Noten den Studierenden fortwährend aktualisierte Informationen über die Chancen, das Studium erfolgreich abschließen zu können. Bei negativen Signalen in Form von schlechten Noten ist es nicht vernünftig, weitere

Investitionen zu tätigen. Je mehr die Erfolgswahrscheinlichkeit unter ein akzeptables Niveau sinkt, desto mehr erscheint ein Abbruch als valide Option.

Das wirft die Frage nach der Rolle der Rahmenbedingungen auf, unter denen die von den Studierenden erbrachten Leistungen benotet werden. Inwieweit Leistung in angemessene Noten umgesetzt werden kann, hängt im Wesentlichen von den verwendeten Benotungsverfahren ab. Diese müssen Leistung erst adäquat erfassen können, und dann in angemessener Weise mit guten bzw. schlechten Noten bewerten. Hier sind aber erhebliche Variationen im Rahmen der studienbegleitenden Prüfungsverfahren zu erwarten – unterschiedliche Prüfungsverfahren bewerten unterschiedliche Aspekte der Leistungserbringung, die „Messgenauigkeit" von Prüfungsverfahren kann stark variieren, und zudem haben die beteiligten Prüfer oft Spielräume bei der Notenvergabe, was einer gewissen Willkür Raum gibt (vgl. Burger und Groß 2014).

Hier kann Unsicherheit darüber entstehen, ob das Leistungsniveau tatsächlich für einen erfolgreichen Studienabschluss ausreicht, weil neben den Anstrengungen der Studierenden Faktoren auf institutioneller Ebene ins Spiel kommen, die außerhalb ihres Einflussbereichs liegen. Jede Prüfungsleistung bietet den Studierenden nicht nur aktualisierte Informationen über ihre Leistungen, sondern auch über die Fairness der Notenvergabe. Wir nehmen an, dass sich die Beeinflussung der Abbruchsintentionen durch die von den Studierenden wahrgenommene Gerechtigkeit sowohl auf indirektem als auch auf direktem Wege vollzieht. Der indirekte Weg geht über eine Beeinflussung der Zufriedenheit mit der akademischen Leistung. Studierende, die die Fairness der Benotungsverfahren kritisch sehen, sind auch weniger zufrieden mit ihren Noten. Dies beeinträchtigt die akademische Integration im Sinne Tintos. Wir erwarten aber auch direkte Effekte der Benotungsverfahren auf Abbruchsintentionen, die über die Leistungszufriedenheit hinausgehen. Je unfairer Benotungsverfahren sind, desto weniger wird Erfolg planbar, so dass Einschätzungen der Erfolgschancen verzerrt werden können. Jedes Mal, wenn unfaire Verfahren zum Einsatz kommen, werden bestehende Unsicherheiten verstärkt, was Abbruchsvorhaben bekräftigt.

Aus Sicht der Studierenden sind zwei Aspekte wichtig, die die Gerechtigkeit von Bewertungsverfahren betreffen, und von denen die Rentabilität der Investitionen in das Studium abhängt: Wie stark ist ihr Einfluss auf den Prozess der Notenvergabe? Und wie sicher können sie sein, dass ihre Leistung sich auch künftig in erfolgsversprechenden Noten niederschlägt? Wir gehen davon aus, dass diese beiden Aspekte der prozeduralen Gerechtigkeit sich zum einen auf die Abbruchsintentionen der Studierenden auswirken und

dass zum anderen das Ausmaß dieser Beeinflussung mit dem soziodemographischen Hintergrund der Studierenden variiert.

4.2.1 Prozedurale Gerechtigkeit

Theorien zu prozeduraler Gerechtigkeit sind als Erweiterung des aus dem Equity-Prinzip abgeleiteten Modells der *distributiven Gerechtigkeit* entstanden (vgl. Leventhal 1980). Distributive Gerechtigkeit fragt danach, inwiefern das *Ergebnis* einer Ressourcenverteilung als gerecht zu werten ist, wobei das Verteilungsergebnis dem Aufwand gegenübergestellt wird, den man betreiben musste, um dieses zu erreichen (vgl. Adams 1965; Homans 1961). Die Fairness der Notenvergabe wurde bislang vorrangig unter solchen distributiven Gesichtspunkten untersucht (vgl. Jasso und Resh 2002; Resh 2009, 2010). Beurteilungen distributiver Gerechtigkeit berücksichtigen aber nicht, unter welchen Umständen das Ergebnis zustande gekommen ist (vgl. Thibaut und Walker 1975). Diese Umstände sind aber wichtig, da faire Verfahren die Wahrscheinlichkeit erhöhen, dass auch das Ergebnis des Verteilungsprozesses fair ist (vgl. Tyler 1994). Die Betrachtung der Verfahren bietet zudem eine *längerfristige* Perspektive: Gerechte Verfahren bieten nicht nur unmittelbar gerechte Ergebnisse, sondern auch Sicherheit darüber, dass das auch in Zukunft so sein wird (vgl. Lind und Tyler 1988).

Was aber kennzeichnet ein Verfahren als fair? Nach Thibaut und Walker (1975) sind das in erster Linie die Mitsprache- und Kontrollmöglichkeiten, die den Empfängern der Ressource bei der Verteilung gewährt werden. Ihre Argumentation zielt darauf ab, dass ein Verfahren eher als fair gesehen wird, wenn das Ergebnis den Empfängern nicht aufgezwungen wurde. Das ist vor allem dann gegeben, wenn die Empfänger selbst in den Verteilungsprozess involviert waren bzw. ihre Ansprüche einbringen konnten. Da sie so selbst einen Teil der Verantwortung für das Ergebnis tragen, wird das Ergebnis eher akzeptiert. Leventhal (1980) erweitert diese auf Einflussmöglichkeiten beschränkte Konzeptionierung später um Objektivität und Genauigkeit der Verteilungsverfahren sowie um die Konsistenz, mit der die Verfahren angewendet werden.

Aufbauend auf die Ansätze von Thibaut und Walker sowie Leventhal findet diese Konzeption prozeduraler Gerechtigkeit bis heute Verwendung (vgl. Cole et al. 2010; Colquitt 2001; Greenberg und Colquitt 2013; van Dijke et al. 2012). Im Unterschied zu der in der empirischen Gerechtigkeitsforschung üblichen Vorgehensweise hat es sich für un-

sere Studie als sinnvoll erwiesen, die prozedurale Gerechtigkeit differenzierter zu betrachten: Wir unterscheiden zwischen einflussbezogener prozeduraler Gerechtigkeit (PGE) und validitätsbezogener prozeduraler Gerechtigkeit (PGV).

4.2.2   Einflussbezogene prozedurale Gerechtigkeit

Das Konstrukt der PGE basiert auf Thibaut und Walkers (1975) Prinzipien der Prozess- und Ergebniskontrolle sowie Leventhals (1980) Regel der Korrigierbarkeit. *Prozesskontrolle* bedeutet, dass die Studierenden Einfluss auf die bei der Benotung eingesetzten Verfahren haben. Das betrifft auch die Frage, welche Aspekte von den Dozentinnen und Dozenten überhaupt als notenrelevant gesehen werden und schließlich in die Note einfließen. Dabei handelt es sich zunächst einmal um die in der Prüfung erbrachte Leistung selbst. Neben dieser kann den Studierenden aber auch die Möglichkeit gegeben werden, ihre eigenen Ansichten in die Notenvergabe einzubringen. Studierende können z.B. in einer Lehrveranstaltung besonderes Engagement gezeigt haben, und in der Folge Anspruch darauf erheben, dass dies auch bei der Benotung gewürdigt wird. Ebenso können Studierende aufgrund ihrer persönlichen Situation (z.B. finanzielle Schwierigkeiten, Krankheit) daran gehindert werden, ihre volle Leistung zu erbringen. Auch das kann bei der Notenvergabe beachtet werden, insofern die Lehrkräfte für solche Einwände offen sind. *Ergebniskontrolle* dagegen bezieht sich nicht auf die bei der Notenfindung verwendeten Kriterien, sondern auf eine direkte Beeinflussung der Note. Zuletzt bedeutet *Korrigierbarkeit* in unserem Kontext, dass den Studierenden erlaubt wird, Widerspruch gegen eine vergebene Note einzulegen.

Wir erwarten, dass die einflussbezogene prozedurale Gerechtigkeit die subjektive Einschätzung des Studienerfolgs beeinflusst. Wenn in Folge „rigider" Benotungsverfahren jegliche Mitsprache verwehrt wird, entsteht leicht der Eindruck, dass man nicht das erhält, was man eigentlich verdient hätte. Durch Einflussmöglichkeiten sind Studierende hingegen nicht einfach passive Empfänger von Noten, sondern sind aktiv an der Bewertung ihrer Leistung beteiligt, wodurch Erfolg grundsätzlich weniger fremdbestimmt zu sein scheint und darüber hinaus sichergestellt werden kann, dass getätigte Investitionen (Vorbereitungszeit, Arbeitsaufwand etc.) auch adäquat honoriert werden. Darum ist es naheliegend, dass eine größere Kontrolle über die Notenvergabe den Studierenden signalisiert, dass ihre Erfolgschancen höher sind.

H1: Je höher Studierende die PGE einschätzen, desto geringer die Intention, das Studium vorzeitig zu beenden.

4.2.3 Validitätsbezogene prozedurale Gerechtigkeit

Wir operationalisieren PGV mittels Leventhals Kriterien der Konsistenz, der Unvoreingenommenheit sowie der Genauigkeit (vgl. Leventhal 1980; siehe auch Colquitt 2001). Die Regel der *Konsistenz* besagt, dass die für eine Ressourcenverteilung verwendeten Bewertungsstandards weder zwischen Personen noch im Zeitverlauf variieren dürfen. Die Dozentinnen und Dozenten müssen demnach bei der Benotung einheitliche Standards verwenden, also ähnliche Leistungen ähnlich bewerten, was die Verteilungsverfahren verlässlich macht. *Unvoreingenommenheit* liegt vor, wenn die verteilende Instanz eine neutrale, unparteiliche Position einnimmt. Dieses Kriterium wird beispielsweise dann verletzt, wenn die Lehrkräfte bestimmte Studierende aufgrund einer persönlichen Abneigung gegen diese schlechter bewerten. Zuletzt ist das Kriterium der *Genauigkeit* zu nennen. Damit eine Verteilungsentscheidung korrekt ausfällt, müssen die verteilenden Instanzen bei der Verteilung alle relevanten Informationen berücksichtigen. In unserem Fall setzt das voraus, dass die Verfahren das Leistungsniveau der Studierenden möglichst genau erfassen.

Wie gut Erfolgschancen eingeschätzt werden hängt davon ab, unter welchen Umständen die Noten zustande kommen. Während hohe PGE Sicherheit bietet, weil die Erfolgschancen aktiv beeinflusst werden können, impliziert hohe PGV, dass die Erfolgschancen nicht durch ein nachlässiges Vorgehen bei der Benotung verschlechtert werden. Wenn Benotungsverfahren hingegen aus Sicht der Studierenden hinsichtlich Konsistenz und Objektivität Mängel aufweisen, entsteht der Eindruck, dass sie ihre Leistung nicht ausreichend in Erfolg umsetzen können. Der Erfolg hängt von Faktoren ab, die außerhalb des Einflussbereichs der Studierenden stehen.

Dabei ist nicht entscheidend, ob Zweifel der Studierenden hinsichtlich der Validität der Verfahren in jedem Fall gerechtfertigt sind. Die Empfänger können unzureichend über das Verfahren informiert sein und darum ein falsches Bild haben (vgl. Roberson und Stewart 2006). So lange das System aber als unfair wahrgenommen wird, wird die Planbarkeit des Studienerfolgs ebenso als eingeschränkt wahrgenommen. Vor einem solchen Hintergrund ist unklar, ob weitere Investitionen in den aktuellen Bildungs-

gang sich auf lange Sicht auszahlen werden. Insofern Studierende die Benotungsverfahren aber so einschätzen, dass ihre Leistung angemessen belohnt wird, wird die Unsicherheit hinsichtlich des Studienerfolgs verringert. Dies wirkt Abbruchsintentionen entgegen.

H2: Je höher Studierende die PGV einschätzen, desto geringer die Intention, das Studium vorzeitig zu beenden.

4.2.4 Herkunftsspezifische Effekte

Es stellt sich nun die Frage, ob der Einfluss prozeduraler Gerechtigkeit auf Abbruchsintentionen für alle Studierenden in gleichem Umfang wirkt, oder ob die Gewichtung, mit welcher die Gerechtigkeitsevaluationen in die Abbruchsintentionen einfließen, gruppenspezifische Differenzen aufweist.

Hohe PGE und PGV bieten den Studierenden Sicherheit darüber, dass ihre Leistung angemessen belohnt wird, und damit auch längerfristige Sicherheit für die Planbarkeit des Studienverlaufs. Das können Studierende bei einer Einschätzung ihrer Investitionsrisiken berücksichtigen. Wenn auch in Zukunft eher mit erfolgreichen Prüfungen zu rechnen ist, weil Erfolg nicht fremdbestimmt ist, können Investitionen in das Studium als rentabel betrachtet werden. Diese Perspektive kommt vor allem solchen Studierenden zu Gute, die aufgrund ihres familiären Hintergrunds eine größere Distanz zum Hochschulsystem haben. Sie sind in der Einschätzung ihrer Erfolgschancen unsicherer und reagieren deshalb stärker auf Faktoren, welche zum einen die Sicherheit des Erfolgs betreffen und zum anderen nicht von den Studierenden selbst beeinflusst werden können (vgl. Bernardi und Cebolla-Boado 2014; Tolsma et al. 2010).

Studierende aus den unteren sozialen Schichten verfügen in geringerem Umfang über kulturelle und soziale Ressourcen, die ihnen die Orientierung und den Umgang mit den Anforderungen im Studium erleichtern könnten (vgl. Bargel und Bargel 2010; Bourdieu und Passeron 1971). Da die Eltern selbst nicht studiert haben, sind ihre Kinder weniger gut auf die Situation im Studium vorbereitet und entsprechend unsicherer. Deshalb nehmen wir an, dass sie der Signalwirkung fairer Benotungsverfahren ein größeres Gewicht geben als Studierende aus den höheren Schichten. Daraus folgt die Erwartung, dass Studierende mit niedrigerem elterlichem sozioökonomischem Status (SES) ihre Abbruchsintentionen stärker von der Fairness der Benotungsverfahren abhängig machen.

H3.1: Je niedriger der elterliche SES der Studierenden ist, desto stärker negativ ist der Zusammenhang zwischen PGE und Abbruchsintentionen.

H3.2: Je niedriger der elterliche SES der Studierenden ist, desto stärker negativ ist der Zusammenhang zwischen PGV und Abbruchsintentionen.

Nachteile von Migranten in den ersten Stufen des Bildungssystems wurden vielfach nachgewiesen (vgl. Becker 2011; Konsortium Bildungsberichterstattung 2006). Bei jenen Migranten, die den Übergang an die Hochschule schaffen, handelt es sich deshalb bezüglich Leistungsfähigkeit und Motivation um eine positiv selektierte Gruppe. Tatsächlich liegt der Anteil an Personen, die sich nach Erwerb der Studienberechtigung für ein Studium entscheiden, unter Migranten höher als unter Nichtmigranten (vgl. Kristen et al. 2008). Doch auch für Studierende mit Migrationshintergrund ist zu erwarten, dass sie mit den Studienbedingungen weniger vertraut und hinsichtlich eines Studienerfolgs unsicherer sind. Auch wenn ihre Eltern im Herkunftsland studiert haben, lassen sich diese Erfahrungen und die damit verbundenen Ressourcen nur bedingt auf Deutschland übertragen (vgl. Friedberg 2000; Kristen und Granato 2007). Ungeachtet der überdurchschnittlichen Motivation der Migranten erscheint es deswegen plausibel, dass auch sie den Gerechtigkeitsevaluationen bei der Einschätzung ihrer Erfolgschancen ein größeres Gewicht zumessen.

Dabei erwarten wir Unterschiede, die sich aus dem Generationenstatus der Studierenden mit Migrationshintergrund ergeben. Die größte Distanz zum Hochschulsystem (und damit die größte Abhängigkeit von der Fairness der Benotungsverfahren) erwarten wir bei Studierenden, die nicht selbst in Deutschland geboren sind (Migranten der 1. Generation). Studierende, die selbst in Deutschland geboren sind, bei denen aber beide Elternteile im Ausland geboren sind (2. Generation), weisen im Vergleich zur ersten Generation geringere Unterschiede zu den Nichtmigranten auf. Schließlich erwarten wir die geringsten Unterschiede zu den Studierenden ohne Migrationshintergrund bei der Gruppe, die selbst in Deutschland geboren ist, aber nur ein Elternteil im Ausland geboren ist (2,5. Generation, vgl. Rumbaut 2004).

H4.1: Je niedriger der Generationenstatus von Studierenden mit Migrationshintergrund, desto stärker ausgeprägt ist der negative Zusammenhang zwischen PGE und Abbruchsintentionen im Vergleich zu Studierenden ohne Migrationshintergrund.

H4.2: Je niedriger der Generationenstatus von Studierenden mit Migrationshintergrund, desto stärker ausgeprägt ist der negative Zusammenhang zwischen PGV und Abbruchsintentionen im Vergleich zu Studierenden ohne Migrationshintergrund.

4.3 Daten und Methode

Zur Prüfung unsere Hypothesen verwenden wir Daten der ersten Welle des *CampusPanels*. Dabei handelt es sich um eine Online-Befragung, die im Wintersemester 2013/14 unter den Studierenden aller Fachbereiche einer der größten deutschen Universitäten durchgeführt wurde. Die Umfrage ist das Ergebnis der Kooperation eines aus vier Teilprojekten bestehenden interdisziplinären Forschungsclusters[2]. Zur Teilnehmerrekrutierung wurden alle Studierenden über den E-Mail-Verteiler der Universität angeschrieben, zusätzlich wurde über Flyer sowie die Social-Media-Kanäle der Universität für die Befragung geworben. Die Teilnehmer hatten die Möglichkeit, am Ende der Umfrage an einer Verlosung teilzunehmen. Insgesamt haben $n = 2564$ Studierende die Umfrage ausgefüllt, was knapp 10% der gesamten Studierendenschaft ausmacht. Die von uns untersuchte Stichprobe umfasst $n = 1393$ Studierende. Die Differenz zum vollständigen erhobenen Sample ergibt sich aus dem Umstand, dass Fragen zur Notenvergabe nur denjenigen Studierenden gestellt werden konnten, welche tatsächlich schon Noten erhalten haben. Die Teilnehmer unseres Samples waren zum Zeitpunkt der Befragung im Schnitt 23,6 Jahre alt und haben 5,5 Semester studiert. Der Frauenanteil liegt mit 62,4% etwas höher als in der Gesamtpopulation der untersuchten Universität (58,3%). Der Anteil der Studierenden mit Migrationshintergrund in unserem Sample beträgt 16,2% (1. Generation: 2,7%; 2. Gen.: 7,1%; 2,5. Gen.: 6,4%). Tabellen mit den Item-Formulierungen aller in Skalen verwendeten Items sowie deren Verteilungen sind in Tab. 4-1 aufgeführt.

---

[2] Nähere Informationen zum CampusPanel finden sich im *CampusPanel User Handbook* (Lang und Hillmert 2014).

Table 4-1 Item-Formulierungen und Verteilungsmaße der in Skalen verwendeten Items

Skala	Mean	SD
Abbruchsintentionen (Cronbachs α = 0,769)		
Ich habe schon öfter daran gedacht, das Studium abzubrechen.	2,68	1,85
Ich denke ernsthaft daran, das Studium ganz aufzugeben.	1,46	1,08
Ich denke ernsthaft daran, mein Hauptfach zu wechseln.	1,51	1,18
Ich werde mein Studium auf jeden Fall bis zum Abschluss weiterführen.	6,49	1,03
Einflussbezogene prozedurale Gerechtigkeit (PGE) (α = 0,691)		
Meine DozentInnen...		
... geben mir die Möglichkeit, meine Ansichten und Empfindungen zur Benotung auszudrücken.	3,29	1,90
... geben mir die Möglichkeit, Einfluss auf die Note zu nehmen.	3,28	1,81
... geben mir die Möglichkeit, gegen die Note Widerspruch einzulegen.	4,19	1,98
Validitätsbezogene prozedurale Gerechtigkeit (PGV) (α = 0,743)		
Meine DozentInnen...		
... wenden die Maßstäbe zur Notenvergabe einheitlich an.	4,77	1,88
... sind bei der Notenvergabe unvoreingenommen.	5,24	1,66
... achten darauf, dass meine Note meinem Kenntnis- und Wissensstand so genau wie möglich entspricht.	4,41	1,71
Leistungszufriedenheit (α = 0,825)		
Meine Leistungen im Studium sind besser, als ich ursprünglich erwartet hatte.	4,24	1,70
Meine Leistungserwartungen und -ansprüche haben sich im Studium voll erfüllt.	4,30	1,55
Mit meiner Studienleistung bin ich zufrieden.	4,86	1,61
Soziale Integration (α = 0,869)		
Mir ist es während meines bisherigen Studiums gut gelungen, Kontakte zu anderen Studierenden aufzubauen.	5,39	1,55
Ich kenne viele Kommiliton(inn)en, mit denen ich mich über fachspezifische Fragen austauschen kann.	5,23	1,57
Ich habe viele Kontakte zu Studierenden aus meinem Semester.	4,85	1,72
Informiertheit vor Studienbeginn (α = 0,767)		
Wie gut war Ihr Informationsstand vor Aufnahme Ihres jetzigen Studiums bezüglich...		
... der Studieninhalte?	4,27	1,66
... der Prüfungsanforderungen?	3,31	1,67
... dem studienbezogenen Zeitaufwand (workload)?	3,70	1,66
Bildungsaspirationen (Einzelitem)		
Der Studienabschluss ist ein wichtiger Zwischenschritt, um meine Lebensziele zu erreichen.	5,91	1,30

4.3.1   Variablen

Für die Messung der *Abbruchsintentionen* verwenden wir vier Items, welche in der *National Educational Panel Study* (NEPS, vgl. Wenzig 2013) verwendet wurden. Beispiel-Items: „Ich habe schon öfter daran gedacht, dass Studium abzubrechen"; „Ich werde mein Studium auf jeden Fall bis zum Abschluss weiterführen". Diese Items wurden anhand einer siebenstufigen Skala abgefragt, von 1 (trifft überhaupt nicht zu) bis 7 (trifft voll und ganz zu) (Cronbachs $\alpha = 0{,}769$). Anhand von konfirmatorischen Faktorenanalysen (CFA) wurden aus den Einzelitems standardisierte Faktorscores für die Abbruchsintentionen berechnet. Diese Scores haben einen Mittelwert von null und eine Standardabweichung von eins. Das bedeutet, dass ein Regressionskoeffizient einer unabhängigen Variable von z.B. -0,5 so zu interpretieren ist, dass bei einer Erhöhung dieser Variable um eine Einheit die vorhergesagten Abbruchsintentionen um eine halbe Standardabweichung verringert werden.

PGE wurde unter Verwendung dreier Items aus dem in der empirischen Gerechtigkeitsforschung verbreiteten Instrument von Colquitt (2001) gemessen. Die Items wurden zunächst ins Deutsche übertragen und anschließend so angepasst, dass der Fokus auf der Notenvergabe durch Dozentinnen und Dozenten liegt. Beispiel-Item: „Meine DozentInnen geben mir die Möglichkeit, Einfluss auf die Note zu nehmen" ($\alpha = 0{,}691$).

PGV wurde ebenfalls anhand des Instruments von Colquitt (2001) gemessen. Beispiel-Item: „Meine DozentInnen sind bei der Notenvergabe unvoreingenommen" ($\alpha = 0{,}743$). Die Items wurden anhand einer siebenstufigen Skala abgefragt, von 1 (trifft überhaupt nicht zu) bis 7 (trifft voll und ganz zu). Anhand von konfirmatorischen Faktorenanalysen wurden aus den jeweiligen Items standardisierte Faktorscores für die beiden Formen der prozeduralen Gerechtigkeit ermittelt.

Der *sozioökonomische Status* des Elternhauses wird über den ISEI-08 (*International Socio-Economic Index of Occupational Status*) in das Modell aufgenommen, welcher Werte zwischen 10 und 90 annehmen kann (vgl. Ganzeboom & Treiman 2014). Dazu wird der jeweils höhere Wert verwendet, wenn Informationen zu beiden Elternteilen vorliegen. Der durchschnittliche ISEI-Score des Untersuchungssamples liegt bei 66,5. 9,7% des Samples haben einen ISEI-Score von unter 40. Um die Interpretation der Interaktionseffekte zu erleichtern, wurde der ISEI-Score um den Mittelwert des Samples zentriert.

Die Variablen zum *Migrationshintergrund* unterscheiden gemäß Hypothesen H4.1 und H4.2 nach dem Generationenstatus. Wenn die Studierenden selbst im Ausland geboren sind, definieren wir sie als Migranten der 1. Generation. Wenn beide Elternteile im Ausland geboren sind, die Person selbst aber in Deutschland, wird sie der 2. Generation zugerechnet. Wenn nur ein Elternteil im Ausland geboren ist und die Person selbst in Deutschland gehört sie dagegen der 2,5. Generation an. Personen, die zwar nicht in Deutschland geboren sind, jedoch vor ihrem sechsten Lebensjahr zugewandert sind (1,75. Generation, vgl. Rumbaut 2004), werden je nach Migrationsstatus der Eltern zur 2. oder 2,5. Generation gezählt, weil sie ihre gesamte Schullaufbahn in Deutschland absolviert haben. Zusätzlich zu den Variablen zum Migrationshintergrund enthält das Modell auch eine Indikator-Variable die angibt, ob die Hochschulzugangsberechtigung vor der Einwanderung erworben wurde (Bildungsausländer). Dadurch wird der Effekt eines Migrationshintergrunds der 1. Generation um etwaige Effekte einer späten Einwanderung bereinigt.

Es ist anzunehmen, dass der *Zufriedenheit mit der akademischen Leistung* eine zentrale Rolle bei der Formierung von Abbruchsentscheidungen zukommt: Studierende, die ihre Leistung als unzureichend einschätzen, werden sich besonders darum sorgen, ob ihre Leistungen für einen Studienabschluss genügen. Andererseits ist davon auszugehen, dass ungerechte Benotungsverfahren die Leistungszufriedenheit der Studierenden beeinträchtigen. Wir erwarten daher, dass die Auswirkungen der Gerechtigkeit auf Abbruchsintentionen teilweise über die Leistungszufriedenheit vermittelt werden: Ungerechte Benotungsverfahren führen dazu, dass die Studierenden mit ihrer Leistung unzufrieden sind, was wiederrum Abbruchsintentionen nahelegt. Um Effekte der Gerechtigkeitsevaluationen identifizieren zu können, die über eine Beeinflussung der Leistungszufriedenheit hinausgehen, muss letztere auch im Modell kontrolliert werden. Zur Modellierung der Leistungszufriedenheit wurden drei Items aus dem NEPS verwendet (vgl. Wenzig 2013). Beispiel: „Meine Leistungserwartungen und -ansprüche haben sich im Studium voll erfüllt" ($\alpha = 0,825$).

Sonstige Kontrollvariablen: Gemäß Tinto (1975) sind Studienabbrüche neben Leistungsaspekten auch durch eine mangelnde soziale Integration in die Hochschule bedingt. Eine entsprechende Kontrollvariable wurde anhand von drei Items gebildet ($\alpha = 0,869$) (NEPS, vgl. Wenzig 2013). Studienabbrüche können neben geringen Erfolgsaussichten auch auf eine falsche Fachwahl zurückgeführt werden (vgl. Heublein et al. 2010). Um das zu berücksichtigen kontrollieren wir den *Informationsstand vor Studienbeginn*

hinsichtlich Studieninhalten, Prüfungsordnung und Arbeitsaufwand. Wir verwenden hier drei Items aus einer Studie von Thiel et al. (2010) (α = 0,767). Um die *Bildungsaspirationen* der Studierenden zu messen wurde gefragt, inwiefern sie der Aussage zustimmen „Der Studienabschluss ist ein wichtiger Zwischenschritt, um meine Lebensziele zu erreichen" (NEPS, vgl. Wenzig 2013).

Die Items zu Leistungszufriedenheit, soziale Integration, Informiertheit vor Studienbeginn sowie zu Bildungsaspirationen wurden jeweils anhand einer siebenstufigen Skala abgefragt. Für Leistungszufriedenheit, soziale Integration und Informiertheit vor Studienbeginn wurden anhand einer CFA aus den Einzelitems standardisierte Faktorscores gebildet.

Die Variable für die *Abiturnote* wurde so codiert, dass entgegen dem deutschen Benotungsschema (1 = gut) ein niedriger Wert eine schlechtere Note darstellt (1 = schlecht). Zusätzlich kontrollieren wir das *Fachsemester* sowie das *Geschlecht* der Studierenden (1 = weiblich, 0 = männlich).

4.3.2 Methode

Wir schätzen die Effekte prozeduraler Gerechtigkeit auf die Abbruchsintentionen der Studierenden, bedingt auf deren soziodemographischen Hintergrund sowie weitere Kontrollvariablen mittels OLS-Regressionen mit robusten Standardfehlern. Neben den zuvor beschriebenen Kontrollvariablen enthalten die Modelle auch Dummy-Variablen zur Kontrolle der Fachzugehörigkeit. Dadurch werden Effekte der prozeduralen Gerechtigkeit um Effekte verschiedener Formen der Studienorganisation bereinigt, die mit den Benotungsverfahren korreliert sein können. Zudem wird vermieden, dass Schätzkoeffizienten aufgrund fachspezifischer Verteilungen der Studierenden auch nach nicht explizit kontrollierten Merkmalen verzerrt sind. Dazu wurden Studiengänge nach Fachgruppen zusammengefasst, wodurch *n* = 20 Fachgruppen unterschieden werden. Die Koeffizienten der Dummies sind aus Platzgründen nicht in den Regressionstabellen enthalten.

4.4 Ergebnisse

Der Aufbau der Analyse gestaltet sich wie folgt: M1 enthält zunächst die Gerechtigkeitsevaluationen sowie die Kontrollvariablen. In M2 wird zusätzlich die Leistungszufriedenheit in das Modell aufgenommen, um zu kontrollieren, dass ein Teil des Effekts der Gerechtigkeitsevaluationen über die Zufriedenheit vermittelt werden kann. M3-M6 betrachten Interaktionen zwischen den beiden Formen der prozeduralen Gerechtigkeit und dem

soziodemographischen Hintergrund der Studierenden, um herkunftsspezifische Effekte der prozeduralen Gerechtigkeit auf Abbruchsintentionen aufzudecken.

4.4.1 Gerechtigkeit und Abbruchsintentionen

Tab. 4-2 zeigt die Ergebnisse einer OLS Regression auf die Abbruchsintentionen der Studierenden. M1 enthält als Prädiktoren die beiden Formen der prozeduralen Gerechtigkeit. Zusätzlich kontrollieren wir die soziodemographischen Variablen, die Bildungsaspirationen, die soziale Integration, die Informiertheit vor Studienbeginn, die Note der Hochschulzugangsberechtigung, das Fachsemester sowie Dummies für die Studienfachgruppe. Die Ergebnisse belegen, dass die Urteile der Studierenden darüber, in welchem Umfang die Benotungsverfahren als gerecht erscheinen, einen signifikanten Einfluss auf das Vorhaben haben, das Studium vorzeitig abzubrechen. Je mehr Einflussmöglichkeiten im Sinne der PGE den Studierenden auf den Prozess der Notenvergabe gewährt werden, desto geringer das Ausmaß, mit dem sie einen Abbruch beabsichtigen. Eine Erhöhung der PGE um eine Standardabweichung verringert die Abbruchsintentionen um 0,069 Standardabweichungen (SD). Erfolg wird besser planbar, wenn die Studierenden ein höheres Maß an Kontrolle über ihre Noten haben.

Table 4-2 Ergebnisse der OLS-Regression zur Erklärung von Abbruchsintentionen

	M1	M2	M3	M4	M5	M6
	b	b	b	b	b	b
PGE	-0,0693**	-0,0281	-0,0295	-0,0275	-0,0128	-0,0286
PGV	-0,1004***	-0,0795**	-0,0769**	-0,0789**	-0,0790**	-0,0597*
SES (ISEI)	-0,0003	-0,0003	-0,0004	-0,0003	-0,0004	-0,0001
Migrationshintergrund *(Ref.: kein Migr.)*						
1. Generation	-0,0195	-0,2720	-0,2892	-0,2721	-0,2752	-0,2778
2. Generation	0,1151	0,0317	0,0370	0,0277	0,0336	0,0206
2,5. Generation	-0,0103	0,0085	0,0085	0,0058	0,0084	0,0071
Leistungszufriedenheit		-0,3223***	-0,3228***	-0,3222***	-0,3232***	-0,3244***
PGE*SES			0,0026*			
PGV*SES				0,0008		
1. Generation*PGE					-0,0539	
2. Generation*PGE					-0,1987*	
2,5. Generation*PGE					-0,0122	
1. Generation*PGV						-0,0427
2. Generation*PGV						-0,1596*
2,5. Generation*PGV						-0,0810
Soziale Integration	-0,1222***	-0,0633*	-0,0603*	-0,0634*	-0,0635*	-0,0631*
Informiertheit	-0,0896***	-0,0508*	-0,0492*	-0,0509*	-0,0488*	-0,0511*
Bildungsaspirationen	-0,1830***	-0,1453***	-0,1462***	-0,1460***	-0,1453***	-0,1454***
Abiturnote	-0,1002*	-0,0450	-0,0442	-0,0447	-0,0471	-0,0466
Abitur im Ausland: ja	0,2050	0,3657	0,3726	0,3692	0,3982	0,3700
Fachsemester	-0,0310***	-0,0286***	-0,0289***	-0,0286***	-0,0286***	-0,0281***
Geschlecht: weiblich	-0,0108	-0,0508	-0,0498	-0,0501	-0,0515	-0,0518
(Nicht abgebildet: Koeffizienten der Fachgruppen-Dummies)						
Konstante	1,5677***	1,1905***	1,1949***	1,1930***	1,1976***	1,1935***
N	1393	1393	1393	1393	1393	1393
Anzahl Parameter	32	33	34	34	36	36
R²	0,154	0,237	0,239	0,237	0,24	0,239

+p< 0,1; *p<0,05; **p<0,01; ***p< 0,001

Das trifft auch auf PGV zu. In H2 wurde die Erwartung formuliert, dass eine als höher wahrgenommene Validität der Benotungsverfahren Abbruchsintentionen verringert. M1 stützt diese Hypothese: Eine Erhöhung der PGV um eine Standardabweichung reduziert Abbruchsintentionen um 0,100 SD. Je weniger aus Sicht der Studierenden die Beziehung zwischen erbrachter Leistung und Note durch eine willkürliche Vorgehensweise bei der

Benotung beeinträchtigt ist, desto klarer können Erfolgschancen eingeschätzt werden, und desto besser ist auf lange Sicht absehbar, ob weitere Investitionen in das Studium sich lohnen werden – dementsprechend wird ein Studienabbruch unwahrscheinlicher.

In M2 wird zusätzlich die Zufriedenheit mit der Leistung im aktuellen Studiengang kontrolliert. Mit jeder Standardabweichung, mit der die Zufriedenheit mit der Leistung größer wird, sinken Abbruchsintentionen um 0,322 SD. Bei schlechter Leistung ist es nicht sinnvoll, weitere Kosten auf sich zu nehmen. Die Ergebnisse in M2 weisen zudem darauf hin, dass Leistungszufriedenheit als Mediator zwischen Gerechtigkeitsevaluationen und Abbruchsintentionen fungiert: PGE lässt unter Kontrolle der Leistungszufriedenheit keinen direkten Effekt mehr erkennen. Studierende, denen größere Einflussmöglichkeiten (PGE) gewährt werden, sind zufriedener mit ihrer Leistung, was das Risiko, über einen Studienabbruch nachzudenken, reduziert – während Unzufriedenheit mit den Leistungen an den Erfolgschancen zweifeln und Abbruchsintentionen stärker werden lässt. Insofern der Effekt der PGE unter der Kontrolle von Leistungszufriedenheit verschwindet, liegt im Falle der PGE eine vollständige Mediation vor. Im Falle der PGV ist die Mediation schwächer ausgeprägt: Der Effekt der PGV wird etwas kleiner, bleibt aber weiterhin hochsignifikant. Eine angemessene Würdigung in adäquaten Benotungsverfahren kann zwar ebenfalls die Zufriedenheit mit der Leistung erhöhen, hat darüber hinaus aber auch einen eigenständigen Effekt auf die Abbruchsintentionen.

4.4.2 Herkunftsspezifische Konsequenzen der Gerechtigkeitsevaluationen

Die Ergebnisse der ersten beiden Modelle sind unter der Annahme zu interpretieren, dass die Effekte additiv sind, so dass alle Studierenden gleichermaßen auf Kontrollmöglichkeiten und valide Verfahren reagieren. Insofern Studierende aus den unteren sozialen Schichten jedoch unsicherer bei der Einschätzung ihrer Erfolgschancen sein können, wurde in H3.1 und H3.2 die Erwartung formuliert, dass die Stärke der Beeinflussung der Abbruchsintentionen durch die Benotungsverfahren bei Studierenden aus statusniedrigen Familien besonders stark ausgeprägt ist. Das wird in M3 und M4 getestet, indem Interaktionseffekte der PGE bzw. PGV mit der sozialen Herkunft der Studierenden in die Regressionsgleichungen aufgenommen werden. M3 zeigt eine auf dem 5%-Niveau signifikante Interaktion zwischen der PGE und dem elterlichen SES. Dieses Ergebnis unterstützt H3.1: Je niedriger die soziale Herkunft der Studierenden, desto stärker hängen deren Abbruchsintentionen von der PGE ab – auch unter Kontrolle der Leistungszufriedenheit. Ein

ähnlicher Zusammenhang war gemäß H3.2 auch zwischen PGV und SES erwartet worden. Die Ergebnisse in M4 liefern jedoch keine Unterstützung für diese Hypothese.
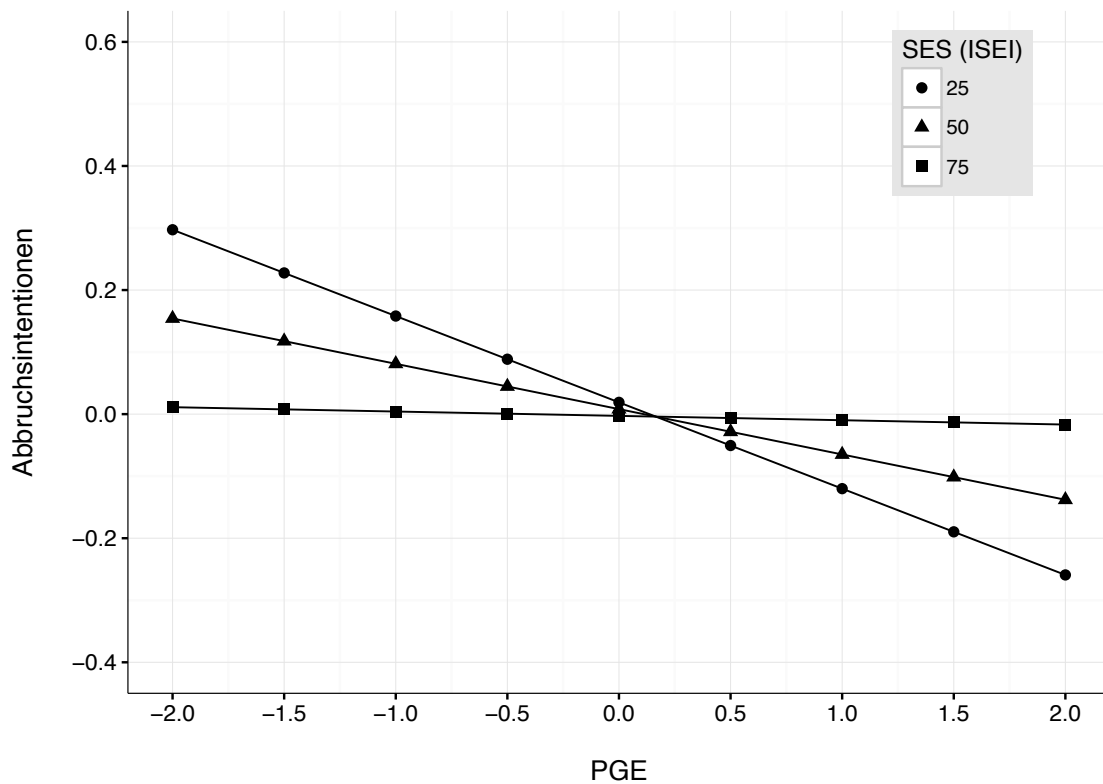
Fig. 4-1 Interaktion PGE*SES. Durchschnittliche Stärke der Abbruchsintentionen entlang des Wertebereichs der PGE für ISEI-Werte von 25, 50 und 75

Mit der signifikanten Interaktion zwischen PGE und SES wird deutlich, dass die oben gezeigte Mediation des Einflusses von PGE auf Abbruchsintentionen durch die Leistungszufriedenheit für die *unteren Statusgruppen nicht vollständig* ist. Das wird in Fig. 4-1 veranschaulicht, in der der Zusammenhang zwischen PGE und Abbruchsintentionen für ISEI-Werte von 25, 50 und 75 separat dargestellt wird. Bei einem ISEI-Wert von 25 (z.B. Imbissverkäufer) verringert eine Erhöhung der PGE um eine Standardabweichung die Abbruchsintention auch unter Kontrolle der Leistungszufriedenheit um 0,139 SD, bei einem ISEI-Wert von 50 (z.B. Fotograf) noch um 0,073 SD. Bei einem ISEI-Wert von 75 (z.B. Ingenieure im Bereich Elektronik) ist nahezu kein Effekt mehr zu sehen. Das bedeutet, dass das Ausmaß der Einflussmöglichkeiten auf unterem SES-Niveau durchaus einen direkten Effekt jenseits der Leistungszufriedenheit auf die Abbruchsintentionen hat, mit steigendem sozialem Status diesen Einfluss aber verliert.
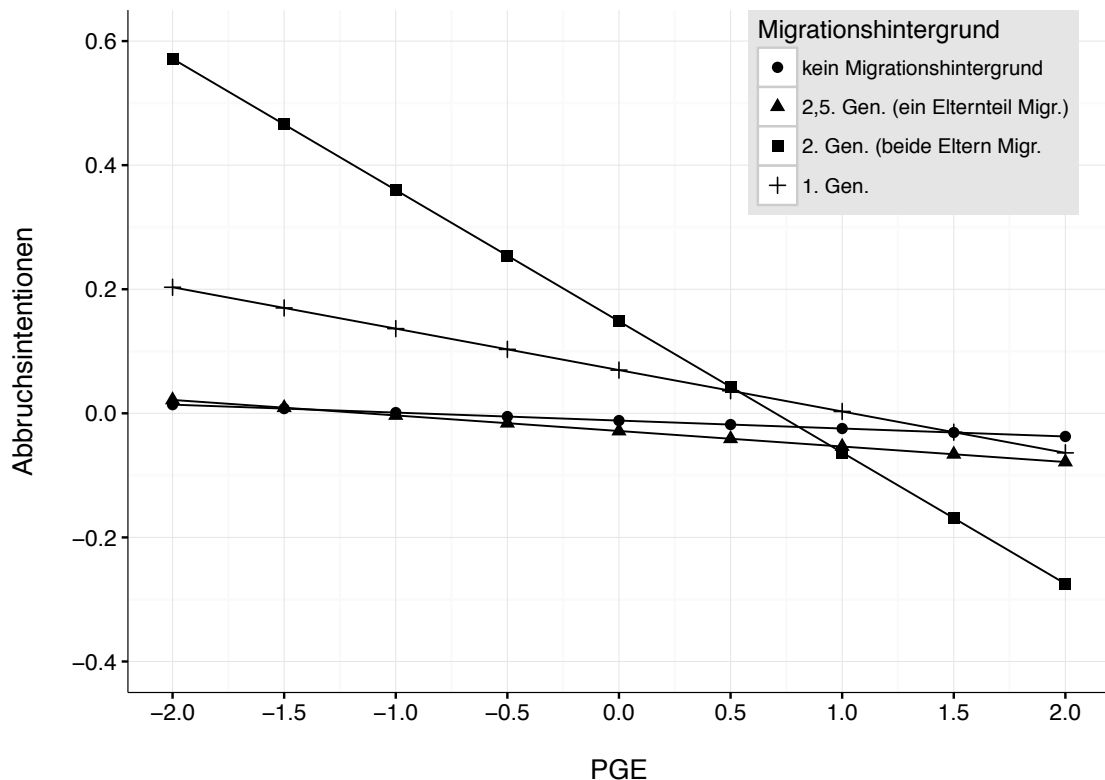
Fig. 4-2 Interaktion PGE*Migrationshintergrund. Nach Migrantengeneration differenzierte durchschnittliche Stärke von Abbruchsintentionen entlang des Wertebereichs der PGE

Die Modelle 1 bis 4 in Tab. 4-2 zeigen für keine der untersuchten Einwanderergenerationen signifikante Unterschiede zu den Abbruchsintentionen einheimischer Studierender. Im Fokus der Hypothese H4.1 steht jedoch die Frage, ob die Abbruchsintentionen von Migranten stärker von der PGE abhängen als die der Nichtmigranten, weil sie – ähnlich wie Studierende mit niedrigem elterlichem SES – nicht auf die Erfahrungen ihrer Eltern mit dem deutschen Hochschulsystem zurückgreifen können. In Folge dessen reagieren sie vermutlich stärker auf Signale, welche die Sicherheit der Einschätzung ihrer Erfolgschancen betreffen. M5 prüft diese These mittels einer Interaktion zwischen PGE und dem Migrationshintergrund, differenziert nach Generationenstatus.

Der in Fig. 4-2 dargestellte Interaktionseffekt stützt unsere Erwartungen. Die PGE zeigt für Nichtmigranten unter Kontrolle der Leistungszufriedenheit keinen direkten Effekt auf die Abbruchsintentionen (b = -0,013), was die Ergebnisse aus M2 widerspiegelt. Für Migranten der zweiten Einwanderergeneration dagegen sehen wir, dass der Effekt der PGE sich signifikant von dem der Nichtmigranten unterscheidet (b = -0,199). Damit vermindert eine Erhöhung der PGE um eine Standardabweichung die Abbruchsintentio-

nen von Migranten der 2. Generation um insgesamt 0,212 SD. Für Migranten der 1. Generation ergibt sich eine Verringerung um insgesamt 0,067 SD. Entgegen H4.1 ist der Effekt der PGE für Migranten der ersten Einwanderergeneration weder stärker als für Migranten der zweiten Generation noch unterscheidet er sich signifikant von dem Effekt für Studierende ohne Migrationshintergrund. Eine mögliche Erklärung hierfür ist, dass Migranten der 1. Generation in unserem Sample nur mit vergleichsweise geringen Fallzahlen vertreten sind ($n = 38$), was die Identifikation signifikanter Effekte erschwert.

Die Regressionsgerade der Migranten der 2,5. Generation verläuft nahezu parallel zu der Gerade der Nichtmigranten. Unter Kontrolle der Leistungszufriedenheit ist damit für beide Gruppen kein direkter Effekt der PGE erkennbar. Insofern laut H4.1 für die 2,5. Generation die geringsten Unterschiede zu Nichtmigranten erwartet wurden, findet diese Hypothese hier Bestätigung. Es scheint, als würde unter Kontrolle der sozialen Lage das Vorhandensein von zumindest einem Elternteil, der das deutsche Schulsystem durchlaufen hat, ausreichen, um etwaige Nachteile auszugleichen.

Fig. 4-3 Interaktion PGV*Migrationshintergrund. Nach Migrantengeneration differenzierte durchschnittliche Stärke von Abbruchsintentionen entlang des Wertebereichs der PGV
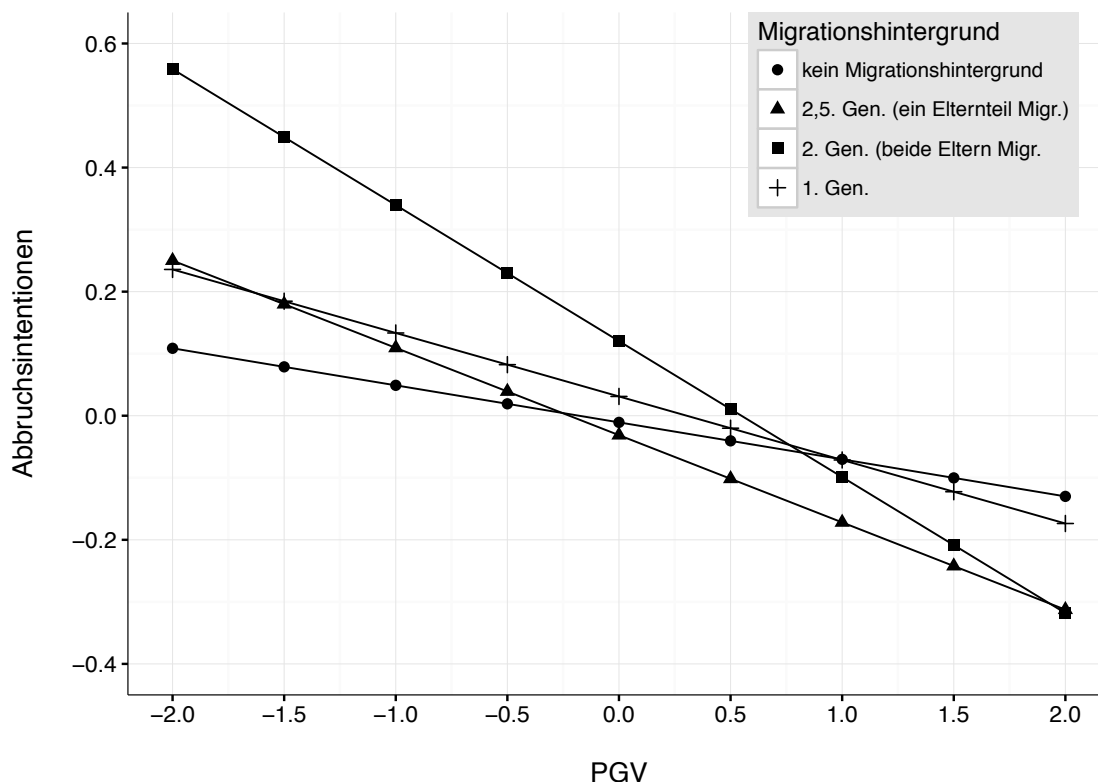
M6 zeigt ein ähnliches Bild: Zwar hat die PGV auch für Nichtmigranten einen signifikanten Effekt auf die Abbruchsintentionen, dieser fällt jedoch deutlich geringer aus als für Studierende mit Migrationshintergrund (siehe Fig. 4-3). Eine Erhöhung der PGV um eine Standardabweichung senkt die Abbruchsintentionen von Nichtmigranten um 0,060 SD, für die 2,5. Generation um 0,141 SD. Die Differenz zwischen beiden Gruppen ist damit vergleichsweise gering und nicht signifikant. Im Gegensatz dazu verläuft die Gerade der 2. Generation wesentlich steiler (-0,219). Wie schon in M5 ist die Differenz der Slopes zwischen Migranten und Nichtmigranten nur für die 2. Generation signifikant. Die Befunde für Migranten der 1. Generation widersprechen unseren Erwartungen[3].

Insgesamt ist festzuhalten, dass Studierende der zweiten Einwanderergeneration besonders stark auf die Fairness der Benotungsverfahren reagieren, wogegen der Unterschied zu den einheimischen Studierenden geringer ausfällt, wenn ein Elternteil in Deutschland geboren ist. Diese Differenzierung stützt das Argument der mangelnden Vertrautheit, weil kaum erkennbare Nachteile vorhanden sind, so lange zumindest ein Elternteil Erfahrungen im deutschen Bildungssystem gemacht hat.

Die Interaktionseffekte zwischen den beiden Formen der prozeduralen Gerechtigkeit und Abbruchsintentionen bestehen auch unter Kontrolle der sozialen Herkunft, und sind folglich nicht dadurch bedingt, dass Migranten im Schnitt häufiger aus statusniedrigen Elternhäusern kommen. Das ist insofern bemerkenswert, als dass sich hier eine Benachteiligung von Studierenden mit Migrationshintergrund abzeichnet, die nicht anhand ihrer sozialen Lage erklärt werden kann. Wie bereits bei der Interaktion zwischen PGE

---

[3] Nach genauer Betrachtung der Modelle konnte unter den Migranten der ersten Einwanderergeneration ein Fall identifiziert werden, der offenbar für Verzerrungen sorgt. Diese Person verfügt über überdurchschnittlich hohe Abbruchsintentionen bei gleichzeitig positiven Bewertungen von PGE und PGV. Wir haben diesen Fall versuchsweise aus unseren Analysen entfernt, mit dem Resultat, dass der Verlauf der Regressionsgeraden in den Interaktionen mit PGE und PGV für Migranten der 1. Generation in die Nähe unserer Erwartungen aus H4.1 und H4.2 rückt (PGE: -0,195 vs. -0,067 in M5; PGV: -0,279 vs. -0,102 in M6). Die Effektgrößen ähneln damit jener der zweiten Einwanderergeneration, die Differenzen zu Nichtmigranten sind aufgrund der kleinen Fallzahl aber weiterhin nicht signifikant. Die Koeffizienten aller anderen Variablen erweisen sich als stabil. Die hohen Abbruchsintentionen des Ausreißers sind durch besonders niedrige Leistungszufriedenheit begründet und erscheinen deshalb plausibel zu sein. Deshalb sehen wir kein hinreichendes Argument für einen grundsätzlichen Ausschluss dieses Falls.

und elterlichem SES sind für Migranten der zweiten Einwanderergeneration erst durch die Berücksichtigung von Interaktionseffekten herkunftsspezifische Differenzen erkennbar. Die Ergebnisse weisen darauf hin, dass Studierende mit Migrationshintergrund ungeachtet ihrer hohen Studienmotivation von widrigen Umständen in Form ungerechter Benotungsverfahren übermäßig verunsichert werden.

4.5 Diskussion

Unsere Ausgangsfrage war, welchen Einfluss die Fairness der Notenvergabe auf die Abbruchsintentionen von Studierenden hat. Um dieser Frage nachzugehen, haben wir auf Basis eines Rational Choice-Frameworks ein theoretisches Modell zur Erklärung der Entstehung von Abbruchsintentionen formuliert. Dieses Modell beinhaltet, dass Studierende sich bei der Einschätzung ihrer Erfolgschancen nicht nur an ihrer Leistung orientieren, sondern auch an der Fairness der Verfahren, anhand welcher ihre Leistung benotet wird.

Wir haben zwei Formen der prozeduralen Gerechtigkeit spezifiziert: Einfluss- und validitätsbezogene prozedurale Gerechtigkeit. PGE betrifft den Umfang, in dem Studierende auf den Benotungsprozess einwirken können. PGV dagegen zielt auf die Frage, inwiefern die zur Benotung eingesetzten Verfahren geeignet sind, um zu validen Ergebnissen zu kommen. Die Ergebnisse unserer empirischen Analysen demonstrieren, dass Abbruchsintentionen von der prozeduralen Gerechtigkeit beeinflusst werden. Die Stärke dieses Zusammenhangs ist nicht für alle Studierenden gleichermaßen ausgeprägt. Mit Studierenden aus statusniedrigen Familien sowie Studierenden mit Migrationshintergrund sind besonders solche Personengruppen betroffen, die an Hochschulen ohnehin schon unterrepräsentiert sind.

Einflussbezogene prozedurale Gerechtigkeit wirkt auf die Abbruchsintentionen, indem über die Möglichkeit zur Einflussnahme auf die Notengebung bessere Noten erzielt werden können, was wiederum die Erfolgschancen verbessert. Unsere Ergebnisse weisen darauf hin, dass Studierende aus statusniedrigen Familien ihre Abbruchsintentionen deutlich stärker von den verfügbaren Einflussmöglichkeiten abhängig machen. Das gilt auch für Migranten, deren Eltern beide nicht in Deutschland geboren wurden. Wir führen die stärkere Auswirkung der PGE auf Abbruchsintentionen bei Studierenden mit niedrigem SES und mit Migrationshintergrund auf die fehlende Vertrautheit mit dem Bereich höherer Bildung zurück sowie mit auf die damit verbundene Überbewertung von Signalen, die sichere Erfolgschancen versprechen. Damit konnten wir einen Mechanismus aufzeigen, welcher herkunftsspezifische Ungleichheit im Bildungserwerb potenziell

verstärken kann, und der bislang in Untersuchungen zu Studienerfolg nicht beachtet wurde.

Ebenso konnte ein Zusammenhang zwischen validitätsbezogener prozeduraler Gerechtigkeit und Abbruchsintentionen gefunden werden. Das Ausmaß der PGV entscheidet über den Umfang, mit dem die erbrachte Leistung in angemessene Noten konvertiert werden kann. Da bei niedriger PGV schlechter eingeschätzt werden kann, ob weitere Investitionen in das Studium einen entsprechenden Return bringen werden, gewinnt ein Abbruch an Attraktivität. Auch hier legen die Ergebnisse den Schluss nahe, dass Studierende, deren Eltern beide im Ausland geboren sind, besonders stark betroffen sind. Sie orientieren sich stärker an Signalen, welche eine sicherere Perspektive bieten. Wenn dagegen nur ein Elternteil im Ausland geboren wurde, sind keine signifikanten Unterschiede zu Nichtmigranten zu beobachten. Dieses Ergebnis betont die Notwendigkeit einer differenzierten Betrachtung des Migrationsstatus, wobei die Interpretation der Befunde für die erste Einwanderergeneration sich aufgrund kleiner Fallzahlen problematisch gestaltet. Eine stärkere Abhängigkeit von der Validität der Benotungsverfahren für Studierende mit niedrigem elterlichem SES konnte entgegen unseren Erwartungen nicht nachgewiesen werden.

Abschließend muss betont werden, dass die hier beschriebenen Zusammenhänge nur an der von uns untersuchten Universität empirisch abgesichert sind. Auch wenn unser theoretisches Grundmodell allgemein formuliert ist, müssen die Annahmen auch in anderen Kontexten überprüft werden. Es ist nicht auszuschließen, dass Merkmale auf Ebene der Universität die Gerechtigkeitsevaluationen sowie die damit verbundenen Reaktionen mitbestimmen. Dazu gehören neben der Zusammensetzung der Studierendenschaft hinsichtlich ihrer sozialen Herkunft auch institutionelle Aspekte wie Studienorganisation und Fachangebot. An der in dieser Studie untersuchten Hochschule werden bspw. keine ingenieurswissenschaftlichen Studiengänge angeboten, womit ein Fachbereich fehlt, in dem Abbrüche überdurchschnittlich oft zu beobachten sind (vgl. Autorengruppe Bildungsberichterstattung 2014). Dazu kommt, dass die Hochschule im Rahmen der Exzellenzinitiative gefördert wird, was sich auf die Bindung der Studierenden an die Universität auswirken kann. Neben anderen Universitäten ist eine Ausweitung der Analysen insbesondere auf Fachhochschulen erstrebenswert, da Studienbedingungen wie auch Abbruchsraten sich an diesen deutlich von denen an Universitäten unterscheiden (vgl. Heublein et al. 2012).

Als weitere Einschränkung ist anzumerken, dass wir aufgrund der Datenlage nur Abbruchsintentionen untersuchen konnten, nicht jedoch tatsächliche Abbrüche. Die Frage nach der Effektstärke der Gerechtigkeitsevaluationen auf Abbrüche kann deshalb vorerst nur mit Vorbehalten beantwortet werden. Eine Diskrepanz zwischen Intention und Abbruch kann beispielsweise dazu führen, dass herkunftsspezifische Unterschiede unterschätzt wurden. So ist es denkbar, dass Studierende aus den oberen Schichten eventuell vorhandene Abbruchsintentionen aufgrund des Statuserhaltsmotivs seltener in die Tat umsetzen.

Zudem sollten Abbrüche künftig auch differenzierter betrachtet werden, da ein Abbruch nicht zwangsweise mit einem Ausscheiden aus dem Bildungssystem gleichzusetzen ist. Vielmehr kann der Abbruch auch in ein neues Studienfach führen, ebenso wie in eine Berufsausbildung. Insbesondere die Unterscheidung zwischen einem Verlassen des Bildungssystems und dem Übergang in einen nicht-tertiären Bildungsweg kann aber anhand der uns zur Verfügung stehenden Daten nicht nachvollzogen werden. Hier bedarf es Längsschnittdaten, über die nicht nur tatsächliche Abbrüche beobachtet werden können, sondern auch das Ziel dieser Abbrüche. Dadurch könnten herkunftsspezifische Effekte besser isoliert werden. Es ist durchaus vorstellbar, dass ein Abbruch für Studierende mit hohem elterlichem SES aufgrund des Statuserhaltsmotivs vorrangig mit einem Fachwechsel einhergeht, während dieser Weg Studierende mit niedrigem elterlichem SES eher in eine Ausbildung führt.

Insgesamt geben die nachgewiesenen herkunftsspezifischen Reaktionen auf wahrgenommene Ungerechtigkeit der Fairness von Benotungsverfahren besondere Relevanz. Sozial ungleiche Bildungsentscheidungen lassen sich an jeder Stelle des Bildungsverlaufs beobachten. Die Studierendenschaft als Ganzes stellt damit bereits eine hochgradig selektive Population dar (vgl. Müller und Pollak 2010). Unsere Ergebnisse legen nun allerdings den Schluss nahe, dass auch im letzten Abschnitt des üblicherweise betrachteten Bildungsverlaufs weitere Hürden vorhanden sind. Studierende aus statusniedrigen Familien und Migranten der 2. Generation orientieren sich bei ihren Abbruchsüberlegungen stärker an den verfügbaren Möglichkeiten zur Beeinflussung der Notenvergabe, letztere sind zudem auch besonders empfindlich bei Verletzungen der validitätsbezogenen prozeduralen Gerechtigkeit. Die besondere Motivation, mit der diese Studierenden in das Studium gehen, kann demzufolge nicht alle herkunftsspezifischen Nachteile ausgleichen.

Dies deutet darauf hin, dass über eine fairere Gestaltung der Notenvergabe eine Reduktion der im Bereich tertiärer Bildung entstehenden sozialen Ungleichheit erzielt

werden kann. Um das bewerkstelligen zu können, muss untersucht werden, wie Gerechtigkeitsurteile zustande kommen. Wir gehen davon aus, dass Gerechtigkeitsevaluationen nicht einfach Ausdruck individueller Präferenzen sind, sondern dass sie vielmehr von der institutionellen, fachspezifischen Umwelt geprägt werden. So wäre etwa zu eruieren, ob eher standardisierte oder eher offene Bewertungsverfahren den Gerechtigkeitsempfindungen der Studierenden entgegenkommen: Hausarbeiten geben den Dozentinnen und Dozenten einerseits mehr Freiräume, die Leistung der Studierenden zu würdigen, andererseits bergen diese Freiräume auch Potential für willkürliches Vorgehen. Klausuren dagegen sind eher standardisiert, was die Verlässlichkeit erhöht und zu einer als höher wahrgenommenen PGV führen sollte. Eine höhere Standardisierung verringert jedoch die Einflussmöglichkeiten, was negative Folgen für die PGE haben kann. Dieses Spannungsverhältnis weist auf einen großen Forschungsbedarf hinsichtlich der Zusammenhänge zwischen Gerechtigkeitsevaluationen und institutionellen Strukturen hin. Unsere Ergebnisse implizieren, dass von einer entsprechenden adäquaten Gestaltung der Bewertungsverfahren vor allem Studierende aus statusniedrigen Familien sowie Studierende mit Migrationshintergrund profitieren würden.

4.6 Literatur

Adams, J. S. (1965). Inequity in social exchange. *Advances in Experimental Social Psychology*, *2*, 267-299.

Autorengruppe Bildungsberichterstattung (2014). *Bildung in Deutschland 2014. Ein indikatorengestützter Bericht mit einer Analyse zur Bildung von Menschen mit Behinderungen*. Bielefeld: Bertelsmann.

Bargel, H., & Bargel, T. (2010). *Ungleichheiten und Benachteiligungen im Hochschulstudium aufgrund der sozialen Herkunft der Studierenden* (S.1-48). Arbeitspapier 202 der Hans-Böckler-Stiftung.

Bean, J. P. (1982). Student attrition, intentions, and confidence: Interaction effects in a path model. *Research in Higher Education*, *17*(4), 291-320.

Becker, R. (2011). Integration von Migranten durch Bildung und Ausbildung - Theoretische Erklärungen und empirische Befunde. In R. Becker (Hrsg.), *Integration durch Bildung. Bildungserwerb von jungen Migranten in Deutschland* (S.11-36). Wiesbaden: VS Verlag für Sozialwissenschaften.

Bernardi, F., & Cebolla-Boado, H. (2013). Previous school results and social background: Compensation and imperfect information in educational transitions. *European Sociological Review, 30*(2), 207-217.

Bourdieu, P., & Passeron, J. C. (1971). *Die Illusion der Chancengleichheit: Untersuchungen zur Soziologie des Bildungswesens am Beispiel Frankreichs*. Stuttgart: Klett.

Breen, R., & Goldthorpe, J. H. (1997). Explaining educational differentials towards a formal rational action theory. *Rationality and Society, 9*(3), 275-305.

Breen, R., van de Werfhorst, H. G., & Jæger, M. M. (2014). Deciding under doubt: A theory of risk aversion, time discounting preferences, and educational decision-making. *European Sociological Review, 30*(2), 258-270.

Burger, R., & Groß, M. (2014). *CampusPanelPaper 3 Gerechtigkeit im Studienverlauf. Ergebnisse der ersten Welle des Campus Panel*. Tübingen: Institut für Soziologie. http://www.campuspanel.ifsoz.uni-tuebingen.de/files/campuspanelpaper3.pdf. Zugegriffen: 08. Dezember 2014.

Cohen-Charash, Y., & Spector, P. E. (2001). The role of justice in organizations: A meta-analysis. *Organizational Behavior and Human Decision Processes, 86*(2), 278-321.

Cole, M. S., Bernerth, J. B., Walter, F., & Holt, D. T. (2010). Organizational justice and individuals' withdrawal: Unlocking the influence of emotional exhaustion. *Journal of Management Studies*, *47*(3), 367-390.

Colquitt, J. A. (2001). On the dimensionality of organizational justice: A construct validation of a measure. *Journal of Applied Psychology, 86(*3), 386-400.

Colquitt, J. A., Scott, B. A., Rodell, J. B., Long, D. M., Zapata, C. P., Conlon, D. E., & Wesson, M. J. (2013). Justice at the millennium, a decade later: A meta-analytic test of social exchange and affect-based perspectives. *Journal of Applied Psychology, 98*(2), 199-236.

Daileyl, R. C., & Kirk, D. J. (1992). Distributive and procedural justice as antecedents of job dissatisfaction and intent to turnover. *Human Relations, 45*(3), 305-317.

Esser, H. (1999). *Soziologie: Spezielle Grundlagen Vol. 1, Situationslogik und Handeln*. Frankfurt: Campus.

Friedberg, R. M. (2000). You can't take it with you? Immigrant Assimilation and the portability of human capital. *Journal of Labor Economics*, *18*(2), 221-51.

Ganzeboom, H. B. G., & Treiman, D. J. (2014). *International Stratification and Mobility File: Conversion Tools*. Amsterdam: Department of Social Research Methodology. http://www.harryganzeboom.nl/ismf/index.htm. Zugegriffen: 20. Februar 2015.

Greenberg, J., & Colquitt, J. A. (Hrsg.) (2013). *Handbook of organizational justice*. Oxford: Psychology Press.

Heublein, U., Schmelzer, R., Sommer, D., & Wank, J. (2012). *Die Entwicklung der Schwund-und Studienabbruchquoten an den deutschen Hochschulen (HIS: Forum Hochschule 3 – 2012)*. Hannover: HIS.

Heublein, U., Hutzsch, C., Schreiber, J., Sommer, D., & Besuch, G. (2010). *Ursachen des Studienabbruchs in Bachelor-und in herkömmlichen Studiengängen (HIS: Forum Hochschule 2 – 2010)*. Hannover: HIS.

Homans, G. C. (1961). *Social behavior: Its elementary forms*. Oxford: Harcourt.

Jasso, G., & Resh, N. (2002). Exploring the sense of justice about grades. *European Sociological Review, 18*(3), 333-351.

Konovsky, M. A., & Cropanzano, R. (1991). Perceived fairness of employee drug testing as a predictor of employee attitudes and job performance. *Journal of Applied Psychology, 76*(5), 698-707.

Konsortium Bildungsberichterstattung (2006). *Bildung in Deutschland. Ein indikatorengestützter Bericht mit einer Analyse zu Bildung und Migration*. Bielefeld: Bertelsmann.

Kristen, C., & Granato, N. (2007). The educational attainment of the second generation in Germany. Social origins and ethnic inequality. *Ethnicities*, *7*(3), 343-366.

Kristen, C., Reimer, D., & Kogan, I. (2008). Higher education entry of Turkish immigrant youth in Germany. *International Journal of Comparative Sociology*, *49*(2-3), 127-151.

Kristen, C. (2014). Migrationsspezifische Ungleichheiten im deutschen Hochschulbereich. *Journal for Educational Research Online, 6*(2), 113-134.

Lang, V., & Hillmert, S. (2014) (Hrsg.). *CampusPanel User Handbook V1.1: Documentation for the Student Panel of the ScienceCampus Tuebingen, wave "a"*. Tübingen: Institut für Soziologie. http://www.campuspanel.ifsoz.uni-tuebingen.de/files/campuspanel_userhandbook_v1_1.pdf. Zugegriffen: 08. Dezember 2014.

Leszczensky, M., Frietsch, R., Gehrke, B., & Helmrich, R. (2009). *Bildung und Qualifikation als Grundlage der technologischen Leistungsfähigkeit Deutschlands: Bericht des Konsortiums „Bildungsindikatoren und technologische Leistungsfähigkeit" (Studien zum deutschen Innovationssystem Nr. 8-2009)*. Berlin: Expertenkommission für Forschung und Innovation.

Leventhal, G. S. (1980). What should be done with equity theory? In K. Gergen, M. Greenberg, & R. Willis (Hrsg.), *Social exchange: Advances in theory and research* (S. 27-55). New York: Plenum Press.

Lind, E. A., & Tyler, T. R. (1988). *The social psychology of procedural justice*. New York: Plenum Press.

McFarlin, D. B., & Sweeney, P. D. (1992). Research notes. Distributive and procedural justice as predictors of satisfaction with personal and organizational outcomes. *Academy of Management Journal, 35*(3), 626-637.

Müller, W., & Pollak, R. (2010). Weshalb gibt es so wenige Arbeiterkinder in Deutschlands Universitäten?. In R. Becker & W. Lauterbach (Hrsg.), *Bildung als Privileg* (4. Aufl., S. 305-344). Wiesbaden: VS Verlag für Sozialwissenschaften.

Resh, N. (2009). Justice in grades allocation: Teachers' perspective. *Social Psychology of Education*, *12*(3), 315-325.

Resh, N. (2010). Sense of justice about grades in school: Is it stratified like academic achievement?. *Social Psychology of Education, 13*(3), 313-329.

Roberson, Q. M., & Stewart, M. M. (2006). Understanding the motivational effects of procedural and informational justice in feedback processes. *British Journal of Psychology*, *97*(3), 281-298.

Rumbaut, R. G. (2004). Ages, life stages, and generational cohorts: Decomposing the immigrant first and second generations in the United States. *International Migration Review*, *38*(3), 1160-1205.

Sarcletti, A., & Müller, S. (2011). Zum Stand der Studienabbruchforschung. Theoretische Perspektiven, zentrale Ergebnisse und methodische Anforderungen an künftige Studien. *Zeitschrift für Bildungsforschung, 1*(3), 235-248.

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

Thibaut, J. W., & Walker, L. (1975). *Procedural justice: A psychological analysis*. Hillsdale, NJ: Erlbaum.

Thiel, F., Blüthmann, I., & Richter, M. (2010). *Ergebnisse der Befragung der Studierenden in den Bachelorstudiengängen an der Freien Universität Berlin.*

Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research, 45*(1), 89-125.

Tinto, V. (1993). *Leaving college: Rethinking the causes and cures of student attrition* (2nd ed.). Chicago: University of Chicago Press.

Tolsma, J., Need, A., & De Jong, U. (2010). Explaining participation differentials in Dutch higher education: The impact of subjective success probabilities on level choice and field choice. *European Sociological Review, 26*(2), 235-252.

Tremblay, M., Cloutier, J., Simard, G., Chênevert, D., & Vandenberghe, C. (2010). The role of HRM practices, procedural justice, organizational support and trust in organizational commitment and in-role and extra-role performance. *The International Journal of Human Resource Management, 21*(3), 405-433.

Tyler, T. R. (1994). Psychological models of the justice motive: Antecedents of distributive and procedural justice. *Journal of Personality and Social Psychology, 67*(5), 850-863.

Van Dijke, M., De Cremer, D., Mayer, D. M., & Van Quaquebeke, N. (2012). When does procedural fairness promote organizational citizenship behavior? Integrating empowering leadership types in relational justice models. *Organizational Behavior and Human Decision Processes, 117*(2), 235-248.

Wenzig, K. (2013): *Startkohorte 5: Studierende (SC5) SUF-Version 3.0.0 Data Manual [Supplement]: Codebook (de), National Educational Panel Study (NEPS)*. Bamberg: Universität Bamberg, Nationales Bildungspanel (NEPS).

5 Study 2: Student perceptions of the fairness of grading procedures: A multilevel investigation of the role of the academic environment

Burger, R. (2016). Student perceptions of the fairness of grading procedures: A multilevel investigation of the role of the academic environment. *Higher Education*. doi:10.1007/s10734-016-0049-1.

Abstract   The purpose of this study is to examine the effects of assessment method (essays vs. exams) and instruction method (seminars vs. lectures) on student perceptions of the fairness of the assessment process. Department-specific combinations of these factors give a unique profile to the assessment process and to the way students interact with faculty. It is argued that the conditions thus created place students in some departments in a more advantageous position when it comes to meeting justice-related expectations. The variables of interest are procedural justice (PJ) and informational justice (IJ). For PJ, aspects regarding the amount of control students can exert on the grading process (PJ-C) are distinguished from aspects regarding the perceived validity of grading procedures (PJ-V). The sample consists of 1,549 students from 48 departments of a German university. Analysis is done via multilevel mixed effects models. Models also check for cross-level interactions between effects of the academic environment and student socioeconomic status (SES). Results show that PJ-C and PJ-V are significantly affected by the assessment method. Higher proportions of essays relative to exams in a department lead to higher ratings of PJ-C, while they decrease ratings of PJ-V. Ratings of IJ are higher as well if assessment is more essay-based, although this only affects low-SES students. Regarding the instruction method, a higher proportion of seminars was found to significantly increase PJ-C and IJ. Again, effects on IJ are moderated by parental SES. Policy implications for reducing feelings of injustice are discussed.

5.1   Introduction

Whether or not the process of grading performance assessments is *fair* is a common concern among university students: Do my professors actually consider the effort I put into this? What is the reasoning behind the grade for my last assignment? Issues of assessment have received considerable attention in the higher education literature. Frequent topics are the shift toward more learner-centered methods (for a recent review, see Pereira et al. 2015) and ways to ensure reliability and validity of grading practices (Bloxham et al. 2015; Chen et al. 2016; Dawson 2015). The significance students attach to a fair assessment process was demonstrated in a recent study by Burger and Groß (2016). They found that students who perceive grading procedures to be more fair are less likely to have dropout intentions. The implications for student retention give special importance to questions regarding the formation of justice evaluations in higher education. The present study approaches this issue from an institutional perspective: How are individual-level justice evaluations affected by institution-level characteristics of university departments? The insights thus gathered can offer guidance for policy measures targeted at a reduction of student feelings of unjust treatment, which in turn could increase retention rates.

The basic theoretical assumptions of this study are derived from the *justice climate* approach. This line of organizational justice research states that justice evaluations cannot be fully understood if they are treated as individual-level phenomena only. Rather, one has to consider that individual experiences and thus justice evaluations are always embedded in a specific social and institutional context (Mossholder et al. 1998; Naumann and Bennett 2000; Whitman et al. 2012). In higher education, university departments provide such a context. Different departments do not simply differ in the subject matter they teach to their students. Rather, they each represent a distinct *academic environment* characterized by specific approaches to teaching, organization of curricula, and methods of assessing student performance (Entwistle and Tait 1990; Neumann 2001; Ramsden 1979). As a result, student experiences vary greatly between departments, which is evident in a number of important outcomes. A large body of literature has shown that the academic environment influences perceptions of the assessment process (Flores et al. 2015; Parpala et al. 2010; Pereira et al. 2016; Sun and Richardson 2016), academic achievement (Brint et al. 2012; Godor 2016; Simpson 2015), as well as student-faculty interactions (Cuseo 2007; Kim and Sax 2014; Severiens and Schmidt 2009).

This raises the question whether the department-specific environment manifests itself in department-specific justice evaluations as well. The idea of justice climates in university departments is supported by a descriptive study by Burger and Groß (2014). They report substantial differences in the average justice evaluations between departments of the same university. This hints at the possibility that the general conditions in some academic environments can be more favorable than in others if the goal is to meet students' justice-related expectations.

This article aims to identify contextual conditions that elicit fairness-related responses and thus lead to the emergence of department-specific justice climates. Multi-level mixed models are used to estimate the effects of department-level predictors on individual-level outcomes in a German university. The focus is on two elements of the academic environment that are central to student experiences: the method of assessing student performance and the method of instruction. Both are subject to significant inter-departmental variation. In the humanities and social sciences, student performance is commonly assessed by means of essay assignments. STEM fields, on the other hand, show a preference for exams such as multiple choice questionnaires (Neumann et al. 2002; Simpson 2015). Likewise, instruction in some programs primarily takes place in large-scale lectures while it is more seminar-based in others.

Department-specific combinations of these institutional characteristics give a unique profile to the assessment process as well as to the way students interact with faculty. As will be argued later on, the conditions thus created can place students in some departments in a more advantageous position when it comes to meeting their justice-related expectations. This is expected to be reflected in student evaluations regarding the fairness of the assessment process.

To provide a more nuanced picture, we also consider that the relationship between academic environment and justice perceptions is not necessarily deterministic. Students are left some leverage when faced with an environment that runs counter to justice-related expectations. It can be argued that how well students make use of these opportunities depends on how well they can adapt to and cope with the affordances of higher education – matters in which students from low socioeconomic status (SES) families were frequently found to have greater problems than their peers from more affluent backgrounds (Ostrove and Long 2007; Pascarella et al. 2004; Rubin 2012; Tinto 1993). Therefore, special attention will be devoted to the situation of low SES students.

5.2 Justice evaluations

The theoretical framework of this study is built on theories of *organizational justice*. Research in this field is concerned with the question of how individuals evaluate the fairness of the allocation of various resources (Greenberg 1990; Greenberg and Colquitt 2005). In this study, the focus is on the procedures used by faculty to assign grades for assessments of student performance. We distinguish *procedural justice* (Leventhal 1980; Thibaut and Walker 1975) and *informational justice* (Greenberg 1993). Procedural justice is further subdivided into aspects regarding the amount of control students can exert on the grading process on the one hand, and aspects regarding the perceived suitability of the procedures to produce valid results on the other (Burger and Groß 2014).

From the point of view of *control-related procedural justice* (*PJ-C*), the grading process appears fair to the students if they are given the possibility to exert influence on this process (Colquitt 2001). PJ-C is based on Thibaut and Walkers's (1975) principles of process control and decision control and Leventhal's (1980) correctability rule. *Process control* means that students have a voice in the grading procedure. This includes the possibility that grading criteria are established in cooperation with the instructor. *Decision control* refers to direct involvement in deciding the grade rather than the grading criteria. The *correctability* rule demands that students can appeal a grade if they feel that the grading decision was flawed. Greater influence relates to justice in that the grading of assessments gives more consideration to the students' needs, thus assigning them a more active role instead of "forcing" grades onto them. As a consequence, students are partially responsible for the result and thus more likely to accept it. However, note that the above principles do not necessarily ensure that procedures are fairer in a sense of being more equitable. Even though decision control can be used to involve students in judging the quality of their own work in a constructive manner, there is a risk of abuse if students simply demand a better grade for no good reason. Students expressing the feeling that they deserve to have some voice in grading decisions can be indicative of a misguided sense of entitlement (Greenberger et al. 2008). Despite that, it is important that fairness is also a matter of perspective. While an impartial observer might consider it fair if a questionable attempt to exert influence is dismissed by the instructor, the student will likely feel injustice as long as the claim was legitimate from *their* point of view.

The concept of *validity-related procedural justice* (*PJ-V*) is derived from Leventhal's (1980) work. According to Leventhal, a distributive procedure is perceived as

fair if the receiving party feels that it is in compliance with certain rules. With regard to the validity of a process, the relevant criteria are bias suppression, consistency, and accuracy (Leventhal 1980). Applied to a higher education context, these rules demand that grading decisions cannot be guided by any partiality for or prejudice against certain students; that the standards used in assigning grades are applied consistently; and that methods are used that are able to accurately capture the students' understanding of the subject.

Informational justice (*IJ*) was introduced as a distinct justice dimension by Greenberg (1993) and describes how individuals are informed about a procedure. Accurate and transparent communication enables individuals to come to a better understanding of the procedures, which in turn increases the likelihood that the procedures themselves are perceived to be fair (Greenberg 1993). In addition, feedback that is reasonable and constructive can be a motivator for improvement in future assignments (Hattie and Timperley 2007). In the present study, a fair communication policy is defined by detailed and thorough explanations on how assessments are graded. Further, explanations and feedback need to be comprehensible; and they have to be communicated in a timely manner (Colquitt 2001).

5.3 University departments, academic environment, and justice climate

For a long time, research in the field of organizational justice has focused on the individual level when explaining the antecedents of justice evaluations. From that point of view, whether or not a procedure is considered to be fair is primarily a reflection of individual preferences and dispositions (Naumann and Bennett 2000). While it is certainly correct that individual-level attributes play an important role, this interpretation does not take into account that procedures are embedded in specific social and institutional contexts. Even though justice evaluations are ultimately expressions of individual sentiment, they are also reactions to actual events and to the conditions surrounding these events (Wegener 1991). A shared institutional context means that individuals are subject to the same conditions, rules, and regulations, and therefore make similar experiences (Mossholder et al. 1998). This promotes the emergence of group-specific *justice climates*, meaning that similarity in experiences and exchange about these experiences will lead to similarities in justice evaluations (Liao and Rupp 2005).

The present article applies this concept to a university setting. Here, the academic environment on the department level describes the basic framework in which learning,

teaching and assessment take place. It is assumed that structural conditions on the institutional level impose a specific form on the assessment process as well as on student-faculty interactions that can potentially impact the fairness of grades from a student perspective. In practice, the size of the impact will depend on a number of factors such as the specific practices and customs within a department (Lindblom-Ylänne et al. 2006; Neumann et al. 2002). There will also be variance due to different approaches chosen by individual instructors (Oleson and Hora 2014; Wilkesmann and Lauer 2015). Nevertheless, the idea here is that the structural conditions in a department give a certain direction to the experiences that students are likely to make. Since students in the same department experience their studies in light of these conditions, it seems reasonable to expect that their justice perceptions would show a certain degree of congruence. At the same time, congruence of justice perceptions within departments points to the possibility that sentiments could differ from students who are exposed to a different environment. The following paragraphs detail the respective roles of assessment method and instruction method in shaping these conditions.

5.3.1 Assessment method

While assessment can take place in a variety of other formats such as peer assessment (Ashenafi 2015; Topping 1998), self-assessment (Orsmond and Merry 2015), and portfolio assessment (Dysthe and Engelsen 2011), this study contrasts *essays* and *exams* as these formats remain by far the most prevalent in the institution studied here. Essays and exams represent two rather different approaches to measuring student performance, which has consequences for student attitudes toward the assessment process (Flores et al. 2015; Maclellan 2001; Scouller 1998). With regard to justice evaluations, the assessment method informs us about the degree to which assessment is standardized. The essay format is on the low end of the standardization spectrum. Unlike exams, it is hardly possible to give an a priori specification of what constitutes a perfect score, and the result cannot always be positively determined to be right or wrong (Norton 1990).

It follows from this that students in departments where assessment is more essay-based should have better opportunities to bring their own perspective into the assessment process and to influence the outcome. The openness of essays also provides that students have better chances to make a compelling argument in the first place - arguing about a wrong answer in a more standardized format is a less promising endeavor. This leads us

to our first hypothesis: A higher proportion of essays in a department is expected to increase ratings of PJ-C (Hypothesis 1_{PJ-C}).

The assessment method also has implications for perceptions of PJ-V. Since exams represent a more standardized approach to performance assessment, they can ensure that criteria like objectivity and consistency are adhered to (Biggs 1973). Essays grant more freedom when judging the results. A positivistic, techno-rationalist conception of assessment as described by Orr (2007) is hardly compatible with essays. Grading decisions are often too complex to be based on a predefined set of universal criteria (Bloxham et al. 2011). This does not mean that the validity of the grading process is necessarily compromised. In fact, one might argue that essays are better suited to capture student understanding of the subject matter (Huang 2016). But due to the lack of standardization, suspicions of arbitrariness are both more likely to arise and harder to dispel. This is complicated by the fact that student views about what is important when judging the quality of an essay can deviate from what faculty are looking for (Norton 1990). Thus, a higher proportion of essays relative to exams is expected to have a negative impact on ratings of PJ-V (Hypothesis 1_{PJ-V}).

With regard to IJ, there is no official policy regarding assessment and feedback on the level of the university studied here. Generally speaking, feedback for exams is usually shorter than for essays. In many cases, feedback for exams is limited to communicating the grade, unless students specifically ask for more information. Yet, feedback for essays can be very sparse as well. Departments work rather autonomously in this regard, and even within departments, there is bound to be variance between individual instructors. Nevertheless, it can be argued that essay-based assessment offers a different platform for the exchange of information between students and faculty. For example, the task assignment for an essay can be discussed in greater detail, which is usually not the case with exams. Since communication channels are established as a by-product of essay-based assessment, the flow of information is facilitated. Therefore, students have more opportunities to satisfy their informational needs, which leads to the hypothesis that ratings of IJ are higher in departments where assessment is more essay-based (Hypothesis 1_{IJ}).

5.3.2   Instruction method

Whether teaching takes place in seminars or in lectures is decisive for how students interact with faculty (Severiens et al. 2015). Interactions in traditional lectures leave students in a passive role. They mostly just follow the instructor's presentation, and apart from their mere presence, their contributions are limited (Severiens and Schmidt 2009). This creates distance between students and faculty, making it harder for the former to actually become active if they need to do so to satisfy their needs (Park and Choi 2014). Research has shown that some of these issues can be compensated by incorporating elements that promote student engagement and interactions with instructors into lectures (Cavanagh 2011; Miller et al. 2013; Roopa et al. 2013). However, interactive elements in lectures are rather uncommon in the university in which data for this study were collected. In stark contrast to traditional lectures, instruction in seminars is more student-centered, encouraging students to actively interact with faculty. This makes it easier for faculty to both recognize and serve student needs, whereas in lectures, it is not uncommon for the majority of students to not have a single direct interaction with faculty in the course of a whole semester (Cuseo 2007).

The mode of interaction implied by the instruction method is assumed to be related to perceptions of PJ-C and IJ. It can be argued that these two justice dimensions are, to some extent, determined by what the students *themselves* make of the situation. With regard to PJ-C, this is obvious: Exerting control on the assessment process is not possible if the students do not act on their own initiative. A professor would not know that a student feels the need or entitlement to influence his or her grades if this is not actively expressed. The same is true for IJ. Even though students can receive assessment-related information without having to become active, this is not always sufficient. The more specific the information a student wants, the less likely that he or she will get this information unless they explicitly ask for it. PJ-V, on the other hand, does not depend on student-faculty interactions: As the demand for impartial and accurate grades does not vary, grading procedures should be valid regardless of whether or not students approach faculty.

In the present study, the curriculum of each department in the sample includes both lectures and seminars. However, the ratio of both formats varies heavily between departments. On the low end of the spectrum, seminars make up less than 40% of classes, while on the upper end, more than 90% of classes take place in seminars. The contrasting properties of seminars and lectures entail that students enrolled in more seminar-based

departments are in a position in which it is more likely that they will become active on their own accord. This creates an environment in which the pursuit of interests regarding PJ-C and IJ is facilitated, which is why a higher proportion of seminars is expected to improve ratings of these justice dimensions (Hypothesis $2_{\text{PJ-C}}$ and Hypothesis 2_{IJ}).

5.3.3 Background-specific moderation of effects of the academic environment

The previous hypotheses were derived under the implicit assumption that students are a homogenous group who respond to their environment in a uniform way. This assumption seems a little too strong, given that justice evaluations do not only tell us something about the structural conditions in the departments, but also about how students experience and interpret these conditions. Thus, students who are exposed to the same conditions can arrive at deviating judgments if the effects of department-level structure are moderated by individual-level attributes. It is expected that the students' socioeconomic background plays such a moderating role. Students from families with low socioeconomic status (SES) were found to have more difficulties adapting to the cultural and social context of university (Ostrove and Long 2007; Pascarella et al. 2004). This includes greater insecurities in interactions with faculty and peers (Bourdieu 1986; Kim and Sax 2009; Tinto 1993). These insecurities have implications for how perceptions of PJ-C and IJ are influenced by the academic environment. Since perceptions of PJ-V are assumed to be unrelated to student-faculty interactions, differential effects on this justice dimension due to insecurities associated with a lower social background are not expected.

Recall that perceptions of PJ-C and IJ are expected to be more negative if the academic environment leaves students in a passive, less involved role; that is, if assessment is exam-based and/or instruction is lecture-based. Yet, even in an unfavorable environment, students can find ways to get what they want. Regarding PJ-C, both exam-based assessment and lecture-based instruction make it harder for students to present an argument for a better grade because of the higher threshold they need to pass to initiate contact and to plead their case. Therefore, it is less likely that they even attempt to do so. Still, a higher threshold does not mean that exerting influence is prohibited per se, it just requires additional efforts. This applies to IJ as well. Even though exams and traditional lectures provide less potential for detailed explanations, this does not mean that information cannot be obtained via other means. Students can still use office hours, email etc., to directly contact faculty to get the information they need.

This requires that students are confident enough in interactions with faculty to contemplate such actions and to show the necessary initiative to follow through. Students from low-SES families are more insecure in these matters than their higher status peers. Thus, they are at a disadvantage when confronted with an academic environment in which the fulfillment of justice-related expectations depends on such requirements. This disadvantage is less pronounced the more interactions with faculty are encouraged by the method of assessment and the method of instruction. It follows from this that the positive effects of a higher proportion of essays and a higher proportion of seminars on both PJ-C and IJ are expected to be larger for students from a lower social background than for those from higher status families (assessment method: Hypothesis 3_{PJ-C} and Hypothesis 3_{IJ}; instruction method: Hypothesis 4_{PJ-C} and Hypothesis 4_{IJ}). Hypotheses are summarized in Table 5-1.

Table 5-1 Research hypotheses

$H1_{PJ-C}$	A higher proportion of essays relative to exams increases ratings of PJ-C.
$H1_{PJ-V}$	A higher proportion of essays relative to exams decreases ratings of PJ-V.
$H1_{IJ}$	A higher proportion of essays relative to exams increases ratings of IJ.
$H2_{PJ-C}$	A higher proportion of seminars relative to lectures increases ratings of PJ-C.
$H2_{IJ}$	A higher proportion of seminars relative to lectures increases ratings of IJ.
$H3_{PJ-C}$	The impact of a higher proportion of essays on ratings of PJ-C is greater for low SES students.
$H3_{IJ}$	The impact of a higher proportion of essays on ratings of IJ is greater for low SES students.
$H4_{PJ-C}$	The impact of a higher proportion of seminars on ratings of PJ-C is greater for low SES students.
$H4_{IJ}$	The impact of a higher proportion of seminars on ratings of IJ is greater for low SES students.

5.4   Data and operationalization

This study uses data from the first wave of the *CampusPanel*. CampusPanel is an online survey that was conducted in the fall semester of 2013 among students of all departments of one of the largest German universities. The data include information on the students' justice evaluations as well as on other study-related attitudes and experiences (Lang and Hillmert 2014). Data on department-level variables were collected from the online course catalog of the university for the semester in which the survey took place. The research sample consists of $n = 1549$ students on the individual level (L1). For some participants, information on ratings of all items that were used as measures of PJ-C and IJ was missing. These cases were excluded from analyses of these justice dimensions. The number of

valid cases is $n = 1496$ for PJ-V and $n = 1530$ for IJ. Descriptive statistics of the study

group are presented in Table 5-2.

Table 5-2 Description of study group

Variable	Mean	SD	Min	Max
Age	23.664	3.89	18	68
SES (ISEI)	66.065	17.395	14	89
No. of semesters studied[1]	5.373	3.334	1	20
Proportion of essays	.312	.162	.066	.767
Proportion of seminars	.65	.243	.053	1.000
Student-faculty ratio	11.96	9.12	1.071	36.634
	Proportion of total			
Gender (female = 1)	.627			
Immigrant (yes = 1)	.175			
Type of degree program				
Undergraduate (Bachelor)	.530			
Graduate (Master)	.134			
State examination (teaching professions)	.151			
State examination (non-teaching professions)	.185			
Faculty				
Economics and social sciences	.194			
Humanities	.266			
Law	.054			
Medicine	.154			
Natural sciences	.309			
Theology	.023			

[1] In German universities, the level of study is expressed in terms of semesters instead of year of study.

Participants are nested in $n = 48$ university departments. All department-level (L2) pre-

dictors are measured on this level. Yet, there is reason to expect additional variance within

departments. This is due to the fact that in German universities, most departments offer

more than one type of degree program (usually Bachelor's degree, Master's degree and

state examination (*Staatsexamen*) for teaching professions). Multiple programs in one

department suggest the possibility that while the aspects of the academic environment

that are measured by the department-level predictors are constant across programs, other,

unobserved factors that could affect justice perceptions may vary (e.g. workload). This

possibility is taken into account when defining the level-2 units for the regression models.

Departments with multiple programs are further subdivided, which leads to a total of $n =$ 93 L2-units.

5.4.1 Instruments

Dependent variables

PJ-C, PJ-V and IJ were each measured using three-item scales. The construction of these scales is based on an instrument developed by Colquitt (2001) that is widely used in organizational justice research. Table 5-3 contains English translations and descriptive statistics of the survey items used in this study. Confirmatory factor analysis (CFA) was used to predict standardized factor scores for the three justice dimensions.

Individual-level predictors

SES: Parental socioeconomic status is measured on a continuous scale using ISEI-08 scores which can take on values between 10 (low status) and 90 (high status) (Ganzeboom and Treiman 2014). If information on both parents was available, the higher value was used. This variable was rescaled so that one unit equals ten ISEI-points.

Additional control variables on the individual level are immigrant background, gender, year of study (measured in number of semesters studied), satisfaction with academic achievement, pre-enrollment information regarding the study program, and digital media use by faculty. Variables for satisfaction with achievement and pre-enrollment information are factor scores derived from a CFA using three individual items, respectively (see Table 5-3).

Table 5-3 Scales, reliabilities, and descriptive statistics

Scale	Mean	SD	Min	Max
PJ-C (Cronbach's α = .677) (adapted from Colquitt 2001)				
My lecturers give me the opportunity to...				
...express my views and feelings regarding the grading process.	3.296	1.898	1	7
...influence the grade.	3.274	1.804	1	7
...appeal grading decisions.	4.193	1.985	1	7
PJ-V (α = .753) (adapted from Colquitt 2001)				
My lecturers...				
...apply grading standards consistently.	4.776	1.871	1	7
...are unbiased when grading.	5.247	1.651	1	7
...make sure that grades reflect my understanding of the subject as good as possible.	4.427	1.704	1	7
IJ (α = .897) (adapted from Colquitt 2001)				
My lecturers...				
...explain thoroughly how grades come about or are justified.	4.451	1.885	1	7
...give explanations regarding the grading process that are reasonable.	4.664	1.760	1	7
...communicate details regarding the grading process in a timely manner.	4.552	1.808	1	7
Satisfaction with academic achievement (α = .798) (adapted from Wenzig 2013)				
My performance is better than I originally expected.	4.246	1.677	1	7
I was able to meet my performance expectations and goals.	4.301	1.544	1	7
I am satisfied with my academic achievement.	4.868	1.591	1	7
Pre-enrollment information regarding program (α = .775) (adapted from Thiel et al. 2010)				
Thinking back on the time before you entered university, how would you rate your knowledgeability regarding the...				
...subject matter	4.286	1.662	1	7
...assessment requirements	3.332	1.669	1	7
...workload	3.689	1.657	1	7
Digital media use by faculty				
My lecturers are open to the use of digital media in the curriculum.	5.290	1.448	1	7

All items were rated on a 7-point scale from "1: strongly disagree" to "7: strongly agree". Translated to English; survey was administered in German.

Department-level predictors

Assessment method: The item measuring the proportion of essays relative to exams in a department is based on CampusPanel data. Participants were asked how many essays and exams they had written thus far. The ratios of these two values were then aggregated to calculate the group mean for each department. Since the data include no details on types of exams students have written, the category "exam" subsumes different formats such as

multiple-choice questionnaires and open-ended questions. Information on the *last* written exam was available for a small subset of the data. Among these, 57.3% stated that their last exam was either partially or fully multiple-choice, so that there seems to be a focus on more standardized instruments.

Instruction method: The proportion of seminars was calculated as the ratio of the total number of classes in each department and the number of classes where instruction takes place in a student-centered format. Variables for proportion of essays and proportion of seminars are scaled so that one unit corresponds to a 10% difference. Additional control variables on the department level are staff-student ratio as well as a categorical variable for type of degree program. The categories are undergraduate (Bachelor), graduate (Master), state examination for teaching professions, and state examination for non-teaching professions.

5.4.2 Method

Multilevel regression models are used to test the hypotheses. By using multilevel modeling, we can account for the fact that students in the same department are subject to the same academic environment. This within-group homogeneity could not be adequately captured if simple OLS-regression is used, which would cause problems with statistical inference, especially in the estimation of standard errors (Raudenbush and Bryk 2002; Snijders and Bosker 2012).

The analytical procedure is divided into several steps. In a multilevel regression model, the intercept term is allowed to vary randomly between departments. As such, we can decompose the total variance in justice judgments into two parts: variance between individuals (L1) and between departments (L2). First, we want to get an estimate of the proportion of the variance that can be attributed to each of the two levels. This is done by fitting a model without predictors (Model 1). Next, a series of regression models is estimated in which each subsequent model adds parameters to the model before it. Model 2 expands on Model 1 by adding individual-level predictors. Model 3 then adds predictors that describe the academic environment on the department level. This gives us information on the effects of assessment method and instruction method on justice perceptions. In addition, we can determine how much of the between-department variance is explained by the department-level predictors. Building on this, Model 4 adds a random slope parameter for the effect of parental SES. This allows us to test whether the effect of SES is

different between departments. Models 5 and 6 then test whether these differences are related to between-department differences in assessment method and instruction method.

Models are estimated using the software Stata 13.1. All continuous predictors are grand mean centered. With regard to the interpretation of the results, dependent variables are standardized factor scores. These scores have a mean of zero and a standard deviation of one. This means that if an independent variable in a regression model shows a regression coefficient of for example -.5, a one-unit increase in this predictor is estimated to lower justice judgments by half a standard deviation (SD).

5.5   Results

5.5.1   Random intercept models with individual-level predictors

As a first step, a model without predictors is fitted for each of the three justice dimensions in order to get estimates of the variance components on the individual and the department level. Model 1 shows an intraclass correlation (ICC) of .088 for PJ-C, .062 for PJ-V and .083 for IJ (see Table 5-4). These values can be interpreted as the proportion of total variance in justice evaluations that is due to variation between departments (L2) as opposed to variation within departments (L1). For example, almost nine percent of the variance in evaluations of PJ-C can be attributed to the institutional context. Likelihood-ratio tests comparing the random intercept models to the results of a pooled OLS-regression without random intercept are significant for all three models ($p < .001$). This means that we can reject the null-hypothesis that the random intercept has zero variance across L2-units for all three justice dimensions, and take this as evidence for the existence of department-specific justice climates.

Model 2 adds L1 predictors to the regressions. This does not provoke any significant change in the L2 intercept variance. The results suggest that the between-department variance found in Model 1 is indeed due to department-level factors rather than a department-specific clustering of students with particular attributes.

Table 5-4 Multilevel regression models, L1 predictors

Level and Variable	PJ-C		PJ-V		IJ	
	M1	M2	M1	M2	M1	M2
Level 1						
Intercept	.125**	.246***	-.019	.054	.070	.199***
	(.044)	(.054)	(.040)	(.053)	(.043)	(.055)
SES (ISEI)		.004		.024+		.013
		(.014)		(.014)		(.014)
Variance components						
Between-dept. (L2) var.	.0865	.0767	.0598	.0631	.0817	.0818
Within-dept. (L1) var.	.8962	.8294	.9065	.8513	.9074	.8438
Additional Information						
ICC	.088		.062		.083	
No. of estimated param.	3	10	3	10	3	10

Standard errors in parentheses
Additional L1 controls: Immigrant background, gender, no. of semesters studied, satisfaction with academic achievement, pre-enrollment information, digital media use by faculty
***$p < .001$; **$p < .01$; *$p < .05$; +$p < .1$

5.5.2   Models with department-level predictors

Variables for proportion of essays, proportion of seminars, and staff-student ratio as well as type of degree program are added in Model 3 (see Table 5-5). Department-level residual variance decreases for all three justice dimensions. For PJ-C, the L2 intercept variance is down to .018 from .087 in Model 1, and for PJ-V it is down to .022 from .060, and for IJ it is down to .044 from .082. This means that 79.3% of the total between-department variance of PJ-C is explained by the department-level predictors; for PJ-V it is 63.3%, and for IJ 46.3%. Thus, the measures of the academic environment explain the majority of the variance in perceptions of PJ-V and PJ-C that occurs between departments, while IJ still shows a significant amount of group-specific variance that is not accounted for by the variables in the model.

Table 5-5 Multilevel regression models, L1 and L2 predictors

Level and Variable	PJ-C M3	PJ-V M3	IJ M3
Level 1			
Intercept	.292***	.167**	.278***
	(.054)	(.056)	(.062)
SES (ISEI)	.005	.024	.015
	(.014)	(.014)	(.014)
Level 2			
Proportion essays	.085***	-.052*	.028
	(.022)	(.024)	(.026)
Proportion seminars	.036*	.025	.044*
	(.017)	(.018)	(.021)
Variance components			
Between-dept. (L2) var.	.0183	.0224	.0444
Within-dept. (L1) var.	.8292	.8576	.8455
Additional Information			
No. of estimated param.	16	16	16

Standard errors in parentheses
Additional L1 controls: Immigrant background, gender, no. of semesters studied, satisfaction with academic achievement, pre-enrollment information, digital media use by faculty
Additional L2 controls: Student-faculty ratio, type of degree program
***$p < .001$; **$p < .01$; *$p < .05$; +$p < .1$

The assessment method shows significant effects on PJ-C and PJ-V. The effect on PJ-C is particularly large: A 10% increase in the proportion of essays relative to exams leads to an average increase in PJ-C of .085 SD. This supports H1$_{PJ-C}$, where it was proposed that essays leave more room for negotiations. If performance was only graded according to predetermined factors such as right or wrong answers, there would be little left to be negotiated. As expected in H1$_{PJ-V}$, the coefficient of proportion of essays shows a negative sign for PJ-V. A 10% increase in the proportion of essays equals a decrease in PJ-V by .052 SD. While essays do leave room for the students' needs to be heard, they also leave room for arbitrariness. Of course, we cannot tell whether or not instructors are actually more likely to grade essays in an arbitrary fashion. Yet, we can tell that students question the validity-related aspects of the grading process to a greater extent than if assessment was more exam-based. Hypothesis H1$_{IJ}$, which stated that the assessment method has an effect on ratings of IJ, is not supported by M3.

The proportion of seminars relative to lectures was predicted to have a significant effect on both PJ-C and IJ. Results are in support of H2$_{PJ-V}$ and H2$_{IJ}$. A 10% increase in

the proportion of seminars increases PJ-C by .036 SD. The effect is slightly more pronounced for IJ, where a 10% increase in seminars equals a .044 SD increase in the dependent variable. Seminars require students to actively participate in direct interaction with faculty. This lowers the threshold students need to overcome to pursue their justice-related needs, which promotes a positive justice climate. As expected, there is no significant effect on PJ-V, which supports the notion that from a student perspective, the use of valid grading criteria does not depend on student-faculty interactions.

Aside from the effects of the academic environment, one particularly interesting finding is the fact that female students give significantly worse ratings in all three justice dimensions (PJ-C: $b = -.272$; $p = .000$; IJ: $b = -.236$; $p = .000$; PJ-V: $b = -.111$; $p = .035$). Indeed, for PJ-C and IJ, gender is one of the most influential predictors in the model. These effects persist even when controlling for the L2 variables, suggesting that the findings are not due to a gender-specific selection into departments in which the academic environment is less favorable for meeting justice-related expectations. Additional research is necessary to find explanations for the large gender gap.

5.5.3   Cross-level interactions

Next, we take a look at the extent to which the effect of the academic environment on PJ-C and IJ is moderated by the students' parental SES. Model 4 (see Table 5-6) adds a random slope for parental SES to the regression equation. This allows effects of SES to vary between L2 units. Likelihood-ratio tests for differences between models with and without the random slope parameter are significant for IJ, but not for PJ-C. This is taken as evidence that the relationship between SES and perceptions of IJ varies between departments. Since this is not the case with PJ-C, cross-level interactions between the L2 predictors and SES are estimated only for IJ. Model 5 provides evidence in support of H3$_{\text{IJ}}$, which stated that the effect of the assessment method on IJ is larger for low-SES students.

Table 5-6 Multilevel regression models, L1 and L2 predictors, random slope and cross-level interactions

| Level and Variable	PJ-C		IJ	
	M4	M4	M5	M6
Level 1				
Intercept	.291***	.263***	.258***	.253***
	(.053)	(.061)	(.061)	(.061)
SES (ISEI)	.011	.010	.013	.012
	(.016)	(.016)	(.016)	(.016)
Level 2				
Proportion essays	.084***	.024	.033	.020
	(.022)	(.025)	(.026)	(.025)
Proportion seminars	.039*	.033+	.031	.049*
	(.017)	(.019)	(.020)	(.020)
Cross-level interaction				
Essays*SES			-.020*	
			(.010)	
Seminars*SES				-.020**
				(.008)
Variance components				
Between-dept. (L2) var.	.0177	.0464	.0473	.0462
Within-dept. (L1) var.	.8229	.8389	.8367	.8355
Slope (L2) var.	.0023	.0023	.0018	.0020
Intercept-slope (L2) cov.	.0021	-.0102	-.0093	-.0097
Additional Information				
No. of estimated param.	18	18	19	19

Standard errors in parentheses
Additional L1 controls: Immigrant background, gender, no. of semesters
studied, satisfaction with academic achievement, pre-enrollment information,
digital media use by faculty
Additional L2 controls: Student-faculty ratio, type of degree program
***$p < .001$; **$p < .01$; *$p < .05$; +$p < .1$

For ease of understanding the interaction between assessment method and SES on IJ, conditional marginal effects are plotted in Fig. 5-1. Here, we see effects of a 10% increase in the proportion of essays on IJ across the range of parental SES as measured by ISEI scores. For students with an ISEI of 25, a 10% increase of the proportion of essays equals an increase of IJ by .114 SD, whereas the same increase in essays increases IJ only by .065 SD if the ISEI is 50. A little further to the right, the 95% confidence bands cross the zero line. This means that the effect ceases to be significant for ISEI scores greater than 52. This suggests that while the means of assessing student performance can influence the way students judge grading-related information policy, this is only true for students from lower status groups. This could explain the lack of a significant main effect of assessment method on IJ we saw in Model 3.
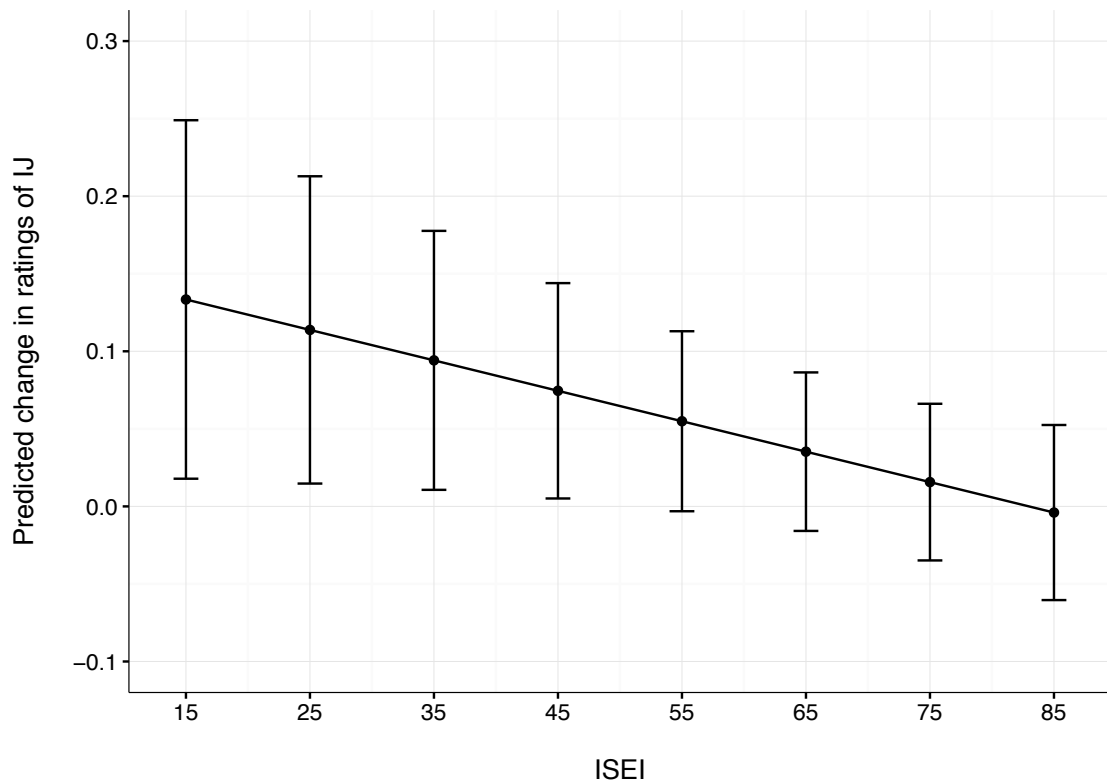
Fig. 5-1 Cross-level interaction assessment method*SES on IJ. Predicted change in ratings of IJ for a 10% increase in the proportion of essays for different levels of SES. 95% confidence bands.

Model 6 shows interactions between instruction method and SES. The effect of proportion of seminars on IJ significantly varies with parental SES (p = .009) and thus provides evidence in favor of H4$_\text{IJ}$. This interaction is visualized in Fig. 5-2. For students with an ISEI of 25, the model predicts ratings of IJ to increase by .13 SD if the proportion of seminars increases by 10%. On the other hand, the same increase in the proportion of seminars improves ratings of IJ only by .08 SD if the ISEI is 50. The effect ceases to be significant for ISEI scores greater than 71.
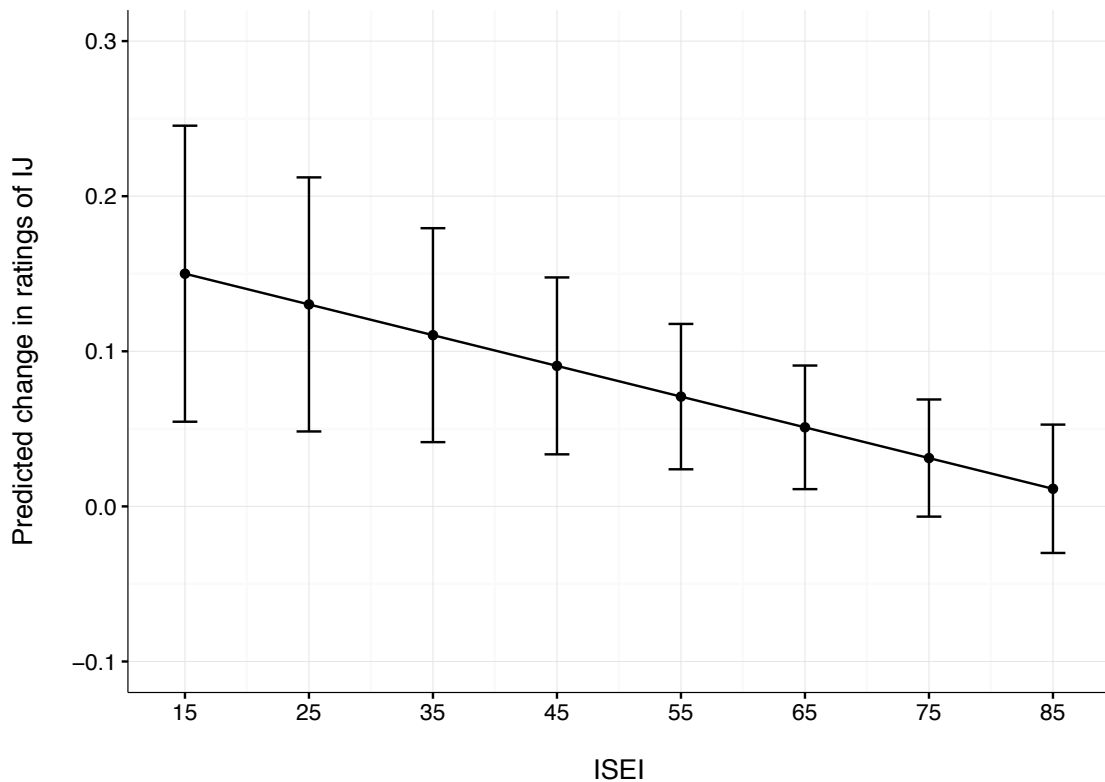
Fig. 5-2 Cross-level interaction instruction method*SES on IJ. Predicted change in ratings of IJ for a 10% increase in the proportion of seminars for different levels of SES. 95% confidence bands.

5.6 Discussion

Results by Burger and Groß (2016) suggest that student retention rates could be improved by reducing feelings of unjust treatment. The purpose of the present study was to contribute to this goal by exploring the role of the academic environment in shaping student perceptions of the fairness of the assessment process. The focus was on the question of how department-specific configurations of assessment method and instruction method influence evaluations of PJ-C, PJ-V, and IJ. With regard to the assessment method, essays were contrasted with exams. In terms of instruction method, this study contrasted seminars and lectures. Results provide evidence for the existence of justice climates within university departments. Ratings of PJ-C were found to be significantly higher in departments were the assessment process is more essay-based as opposed to exam-based. On the other hand, it could be demonstrated that essay-based assessment has a detrimental effect on perceptions of PJ-V. Perceptions of IJ are positively correlated with a larger proportion of essays; effects were found to be especially strong for students from low status families. As for the method of instruction, a more student-centered approach to

teaching in form of seminars as opposed to lectures proved to be beneficial for ratings of PJ-C and IJ. Again, the effect on IJ is moderated by parental SES.

Taken together, these results deliver valuable insights into how individual justice evaluations depend on the academic environment. Whether or not the grading process appears fair is to some extent a matter of being enrolled in the right (or wrong) department. Inter-departmental variation in assessment method and instruction method can lead to substantial differences in justice climates. This can be demonstrated by looking at the departments in our sample. The proportion of essays is below 11.4% in one tenth of departments, while it is below 64.9% in 90% of departments. By plugging these values into the regression equation for M3, we get an average difference in ratings of PJ-C of .453 SD between departments in the first and the last decile. Therefore, students in the first group appear to be at a major disadvantage when it comes to exerting influence on the grading process. Conversely, grading of essay-based assessment can lack transparency and at worst appear arbitrary when compared to a more standardized approach. Again comparing departments in the first and last decile, our model predicts PJ-V to be rated lower by .278 SD by students in departments with the highest proportion of essays. Regarding the instruction method, the proportion of seminars is below 39.7% in one tenth of departments in the sample while it is over 94.7% in the highest decile. The model predicts an average difference in perceptions of PJ-C of .200 SD between departments in the first and the last decile; .244 SD for IJ. Thus, we can see obvious benefits for students in departments where instruction is primarily seminar-based.

The significant cross-level interactions add another dimension to these results: Since low-SES students' perceptions of IJ exhibit a stronger dependence on the academic environment, inter-departmental differences are even more pronounced for this group. For students with an ISEI of 25, ratings of IJ are predicted to differ by .609 SD between departments in the first and the last decile of proportion of essays and by .716 SD when looking at the proportion of seminars. These comparisons point to structural inequalities between university departments that are usually overlooked. Given the particularly delicate standing of low-SES students in higher education, the findings in this study underline the importance of creating an environment that considers the needs of this group.

The conclusions have implications for policy decision making. From the point of view of PJ-C and IJ, it is tempting to recommend an increase in the proportion of essays to improve justice climate. Unfortunately, the negative effects of essays on perceptions of PJ-V lead to an obvious dilemma. Still, an argument in favor of essays can be made.

While we do know that essays are more likely to create an impression of questionable grading practices, we do not know the extent to which these concerns are grounded. That is, would ratings of PJ-V still be lower in essay-based assessment if students had sufficient information on the grading process? Since grading criteria for essays are more complex, vague explanations given for an essay could be perceived more negatively than vague explanations given for an exam. It can be argued that one of the best ways to mitigate validity-related concerns is to make the process transparent through quality feedback (Carless 2006; Lizzio and Wilson 2008). Therefore, attempts to reduce feelings of injustice by prioritizing essay-based assessment can only be successful if they are accompanied by measures that ensure that a certain level of feedback quality and transparency is maintained. Such policies are still not present in many universities, including the one where this study took place. Regarding the instruction method, the case is rather straightforward: Seminar-based instruction facilitates student-faculty interaction and should thus be preferred when compared to traditional lectures. This adds to the existing literature in favor of a more student-centered approach to teaching (Cuseo 2007; Severiens and Schmidt 2009). Note, however, that these findings are based on data from a single German university. Therefore, the applicability of these recommendations in other contexts needs to be substantiated by further research.

Of course, these policy recommendations need to be measured against what could realistically be implemented given the resources available to a particular institution. Large scale lectures enable schools to teach quantities of students that would otherwise exceed available capacities (Maringe and Sing 2014). Likewise, the benefits of essays over exams come at the cost of increasing the workload for faculty when grading assignments – let alone the efforts that are necessary to provide extensive feedback (Price et al. 2010). Thus, there is bound to be friction between a study organization that is in line with students' justice-related expectations and one that is feasible in light of limited resources and other situational constraints. Decisions need to be made by evaluating the status quo on a case-by-case basis: As long as the choice of assessment method and instruction method is not dictated by the circumstances, there is a case for the option that is most beneficial for a positive justice climate.

In closing this article, it is necessary to point out some limitations. First, given that the sample is comprised of students of a single German university, it is difficult to assess the extent to which the results can be generalized. It can be argued that the theoretical mechanisms that were proposed in this study are general enough to not be limited

to the context of the study. Yet, it is necessary to investigate how effects of the academic environment manifest themselves in different institutions and for different student populations, especially in cross-country comparisons. This is particularly important in terms of the policy recommendations derived from this study.

Next, future research should also consider the effects of a more varied set of methods of assessment and instruction by using more fine-grained data. Essays, exams, seminars, and lectures are rather broad categories that subsume a range of approaches to assessment and teaching. Assessment via exam can take place in many different formats, from multiple-choice questionnaires to open-ended questions. One could argue that negative effects of the exam format on perceptions of PJ-C should be lower in case of open questions than for multiple-choice. Likewise, advantages of exams in terms of PJ-V could be less pronounced for open formats when compared to a tightly structured approach. In the same vein, a binary representation of instruction method cannot account for the possibility that some lectures can engage students in interactive processes. Therefore, differentiated analyses are necessary to provide a more accurate picture of student perceptions.

Finally, this study was focused on structural characteristics of departments. Although the structure specifies the general direction of assessment and student-faculty interactions, it does not fully determine the outcome. Faculty actions are also guided by disciplinary norms and department-specific customs (Becher 1989; Biglan 1973; Neumann et al. 2002). This can counteract the tendencies suggested by the structural conditions. For example, while we found considerable between-department variance in evaluations of IJ, the larger part of these differences could not be explained by our department-level predictors. A possible explanation for this could be that even though the academic environment defines the framework for the transmission of information, the *actual* feedback practices might depend on disciplinary customs as well. We cannot rule out the possibility that some disciplines simply assign little value to extensive feedback, thus canceling out the benefits of an environment that should otherwise be favorable for perceptions of IJ. Likewise, essays might offer better opportunities to influence the grade, but this is of little use if this type of student involvement is generally frowned upon in a department. Therefore, it is advised that future research should also consider faculty perceptions of these matters. This would provide a broader and more differentiated knowledge base for policies that aim to reduce feelings of injustice and to close interdepartmental gaps in justice climates.

5.7   References

Ashenafi, M. M. (2015). Peer-assessment in higher education–twenty-first century practices, challenges and the way forward. *Assessment & Evaluation in Higher Education*. doi:10.1080/02602938.2015.1100711.

Becher, T. (1989). *Academic tribes and territories: Intellectual enquiry and the cultures of disciplines*. Milton Keynes: SRHE/Open University Press.

Biggs, J. B. (1973). Study behaviour and performance in objective and essay formats. *Australian Journal of Education, 17*(2), 157-167.

Biglan, A. (1973). The characteristics of subject matter in different academic areas. *Journal of Applied Psychology, 57*(3), 195-203.

Bloxham, S., Boyd, P., & Orr, S. (2011). Mark my words: The role of assessment criteria in UK higher education grading practices. *Studies in Higher Education, 36*(6), 655-670.

Bloxham, S., Hughes, C., & Adie, L. (2015). What's the point of moderation? A discussion of the purposes achieved through contemporary moderation practices. *Assessment & Evaluation in Higher Education*. doi:10.1080/02602938.2015.1039932.

Bourdieu, P. (1986). The forms of capital. In J. G. Richardson (Ed.), *Handbook of theory and research for the sociology of education* (pp. 241-258). New York: Greenwood.

Brint, S., Cantwell, A. M., & Saxena, P. (2012). Disciplinary categories, majors, and undergraduate academic experiences: Rethinking Bok's "underachieving colleges" thesis. *Research in Higher Education, 53*(1), 1-25.

Burger, R., & Groß, M. (2014). *CampusPanelPaper 3 Gerechtigkeit im Studienverlauf. Ergebnisse der ersten Welle des Campus Panel*. Tübingen: Institut für Soziologie. http://www.campuspanel.ifsoz.uni-tuebingen.de/files/campuspanelpaper3.pdf. Accessed 10 November 2015.

Burger, R., & Groß, M. (2016). Gerechtigkeit und Studienabbruch. Die Rolle der wahrgenommenen Fairness von Benotungsverfahren bei der Entstehung von Abbruchsintentionen. *Zeitschrift für Erziehungswissenschaft, 19*(3), 625-647. doi:10.1007/s11618-016-0672-8.

Carless, D. (2006). Differing perceptions in the feedback process. *Studies in Higher Education*, *31*(2), 219-233.

Cavanagh, M. (2011). Students' experiences of active engagement through cooperative learning activities in lectures. *Active Learning in Higher Education*, *12*(1), 23-33.

Chen, C. Y., Chang, H., Hsu, W. C., & Sheen, G. J. (2016). Learning, behaviour and reaction framework: A model for training raters to improve assessment quality. *Assessment & Evaluation in Higher Education*. doi:10.1080/02602938.2016.1180663.

Colquitt, J. A. (2001). On the dimensionality of organizational justice: A construct validation of a measure. *Journal of Applied Psychology*, *86*(3), 386-400.

Cuseo, J. (2007). The empirical case against large class size: Adverse effects on the teaching, learning, and retention of first-year students. *The Journal of Faculty Development*, *21*(1), 5-21.

Dawson, P. (2015). Assessment rubrics: Towards clearer and more replicable design, research and practice. *Assessment & Evaluation in Higher Education*. doi:10.1080/02602938.2015.1111294.

Dysthe, O., & Engelsen, K. S. (2011). Portfolio practices in higher education in Norway in an international perspective: Macro-, meso- and micro-level influences. *Assessment & Evaluation in Higher Education*, *36*(1), 63-79.

Entwistle, N., & Tait, H. (1990). Approaches to learning, evaluations of teaching, and preferences for contrasting academic environments. *Higher Education*, *19*(2), 169-194.

Flores, M. A., Veiga Simão, A. M., Barros, A., & Pereira, D. (2015). Perceptions of effectiveness, fairness and feedback of assessment methods: A study in higher education. *Studies in Higher Education*, *40*(9), 1523-1534.

Ganzeboom, H. B. G., & Treiman, D. J. (2014). *International Stratification and Mobility File: Conversion Tools*. Amsterdam: Department of Social Research Methodology. http://www.harryganzeboom.nl/ismf/index.htm. Accessed 10 November 2015.

Godor, B. P. (2016). Revisiting differential grading standards anno 2014: An exploration in Dutch higher education. *Assessment & Evaluation in Higher Education*. doi:10.1080/02602938.2016.1173186.

Greenberg, J. (1990). Organizational justice: Yesterday, today, and tomorrow. *Journal of Management, 16*(2), 399-432.

Greenberg, J. (1993). The social side of fairness: Interpersonal and informational classes of organizational justice. In R. Cropanzano (Ed.), *Justice in the workplace: Approaching fairness in human resource management* (pp. 79-103). Hillsdale, NJ: Erlbaum.

Greenberg, J., & Colquitt, J. A. (Eds.). (2005). *Handbook of organizational justice*. Mahwah, NJ: Erlbaum.

Greenberger, E., Lessard, J., Chen, C., & Farruggia, S. P. (2008). Self-entitled college students: Contributions of personality, parenting, and motivational factors. *Journal of Youth and Adolescence, 37*(10), 1193-1204.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112.

Huang, V. (2016). An Australian study comparing the use of multiple-choice questionnaires with assignments as interim, summative law school assessment. *Assessment & Evaluation in Higher Education*. doi:10.1080/02602938.2016.1170761.

Kim, Y. K., & Sax, L. J. (2009). Student–faculty interaction in research universities: Differences by student gender, race, social class, and first-generation status. *Research in Higher Education, 50*(5), 437-459.

Kim, Y. K., & Sax, L. J. (2014). The effects of student–faculty Interaction on academic self-concept: Does academic major matter?. *Research in Higher Education, 55*(8), 780-809.

Lang, V., & Hillmert, S. (2014) (Eds.). *CampusPanel User Handbook V1.1: Documentation for the Student Panel of the ScienceCampus Tuebingen, wave "a"*. Tübingen: Institut für Soziologie. http://www.campuspanel.ifsoz.uni-tuebingen.de/files/campuspanel_userhandbook_v1_1.pdf. Accessed 10 November 2015.

Leventhal, G. S. (1980). What should be done with equity theory? In K. Gergen, M. Greenberg, & R. Willis. (Eds.), *Social exchange: Advances in theory and research* (pp. 27-55). New York: Plenum Press.

Liao, H., & Rupp, D. E. (2005). The impact of justice climate and justice orientation on work outcomes: A cross-level multifoci framework. *Journal of Applied Psychology*, *90*(2), 242-256.

Lindblom-Ylänne, S., Trigwell, K., Nevgi, A., & Ashwin, P. (2006). How approaches to teaching are affected by discipline and teaching context. *Studies in Higher Education*, *31*(3), 285-298.

Lizzio, A., & Wilson, K. (2008). Feedback on assessment: Students' perceptions of quality and effectiveness. *Assessment & Evaluation in Higher Education*, *33*(3), 263-275.

Lizzio, A., Wilson, K., & Simons, R. (2002). University students' perceptions of the learning environment and academic outcomes: Implications for theory and practice. *Studies in Higher Education*, *27*(1), 27-52.

Maclellan, E. (2001). Assessment for learning: The differing perceptions of tutors and students. *Assessment & Evaluation in Higher Education*, *26*(4), 307-318.

Maringe, F., & Sing, N. (2014). Teaching large classes in an increasingly internationalising higher education environment: Pedagogical, quality and equity issues. *Higher Education*, *67*(6), 761-782.

Miller, C. J., McNear, J., & Metz, M. J. (2013). A comparison of traditional and engaging lecture methods in a large, professional-level course. *Advances in Physiology Education*, *37*(4), 347-355.

Mossholder, K. W., Bennett, N., & Martin, C. (1998). A multilevel analysis of procedural justice context. *Journal of Organizational Behavior*, *19*(2), 131-141.

Naumann, S. E., & Bennett, N. (2000). A case for procedural justice climate: Development and test of a multilevel model. *Academy of Management Journal*, *43*(5), 881-889.

Neumann, R. (2001). Disciplinary differences and university teaching. *Studies in Higher Education*, *26*(2), 135-146.

Neumann, R., Parry, S., & Becher, T. (2002). Teaching and learning in their disciplinary contexts: A conceptual analysis. *Studies in Higher Education*, *27*(4), 405-417.

Norton, L. S. (1990). Essay-writing: What really counts?. *Higher Education*, *20*(4), 411-442.

Oleson, A., & Hora, M. T. (2014). Teaching the way they were taught? Revisiting the sources of teaching knowledge and the role of prior experience in shaping faculty teaching practices. *Higher Education*, *68*(1), 29-45.

Orr, S. (2007). Assessment moderation: Constructing the marks and constructing the students. *Assessment & Evaluation in Higher Education*, *32*(6), 645-656.

Orsmond, P., & Merry, S. (2015). Tutors' assessment practices and students' situated learning in higher education: Chalk and cheese. *Assessment & Evaluation in Higher Education*. doi:10.1080/02602938.2015.1103366.

Ostrove, J. M., & Long, S. M. (2007). Social class and belonging: Implications for college adjustment. *The Review of Higher Education*, *30*(4), 363-389.

Park, E. L., & Choi, B. K. (2014). Transformation of classroom spaces: Traditional versus active learning classroom in colleges. *Higher Education*, *68*(5), 749-771.

Parpala, A., Lindblom-Ylänne, S., Komulainen, E., Litmanen, T., & Hirsto, L. (2010). Students' approaches to learning and their experiences of the teaching–learning environment in different disciplines. *British Journal of Educational Psychology*, *80*(2), 269-282.

Pascarella, E. T., Pierson, C. T., Wolniak, G. C., & Terenzini, P. T. (2004). First-generation college students: Additional evidence on college experiences and outcomes. *Journal of Higher Education*, *75*(3), 249-284.

Pereira, D., Flores, M. A., & Niklasson, L. (2015). Assessment revisited: A review of research in Assessment and Evaluation in Higher Education. *Assessment & Evaluation in Higher Education*. doi:10.1080/02602938.2015.1055233.

Pereira, D., Niklasson, L., & Flores, M. A. (2016). Students' perceptions of assessment: A comparative analysis between Portugal and Sweden. *Higher Education*. doi:10.1007/s10734-016-0005-0.

Price, M., Handley, K., Millar, J., & O'Donovan, B. (2010). Feedback: All that effort, but what is the effect?. *Assessment & Evaluation in Higher Education*, *35*(3), 277-289.

Ramsden, P. (1979). Student learning and perceptions of the academic environment. *Higher Education*, *8*(4), 411-427.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). Los Angeles, CA: Sage.

Roopa, S., Bagavad Geetha, M., Rani, A., & Chacko, T. (2013). What type of lectures students want? - A reaction evaluation of dental students. *Journal of Clinical and Diagnostic Research*, *7*(10), 2244-2246.

Rubin, M. (2012). Social class differences in social integration among students in higher education: A meta-analysis and recommendations for future research. *Journal of Diversity in Higher Education*, *5*(1), 22-38.

Scouller, K. (1998). The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay. *Higher Education*, *35*(4), 453-472.

Severiens, S. E., & Schmidt, H. G. (2009). Academic and social integration and study progress in problem based learning. *Higher Education*, *58*(1), 59-69.

Severiens, S. E., Meeuwisse, M., & Born, M. (2015). Student experience and academic success: Comparing a student-centred and a lecture-based course programme. *Higher Education*, *70*(1) 1-17.

Simpson, A. (2015). Assessment and its outcomes: The influence of disciplines and institutions. *Assessment & Evaluation in Higher Education*. doi:10.1080/02602938.2015.1052369.

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). Los Angeles, CA: Sage.

Sun, H., & Richardson, J. T. (2016). Students' perceptions of the academic environment and approaches to studying in British postgraduate business education. *Assessment & Evaluation in Higher Education*, *41*(3), 384-399.

Thibaut, J. W., & Walker, L. (1975). *Procedural justice: A psychological analysis*. Hillsdale, NJ: Erlbaum.

Thiel, F., Blüthmann, I., & Richter, M. (2010). *Ergebnisse der Befragung der Studierenden in den Bachelorstudiengängen an der Freien Universität Berlin*.

Tinto, V. (1993). *Leaving college: Rethinking the causes and cures of student attrition* (2nd ed.). Chicago: University of Chicago Press.

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, *68*(3), 249-276.

Wegener, B. (1991). Relative deprivation and social mobility: Structural constraints on distributive justice judgments. *European Sociological Review*, *7*(1), 3-18.

Wenzig, K. (2013): *Startkohorte 5: Studierende (SC5) SUF-Version 3.0.0 Data Manual [Supplement]: Codebook (de), National Educational Panel Study (NEPS)*. Bamberg: Universität Bamberg, Nationales Bildungspanel (NEPS).

Whitman, D. S., Caleo, S., Carpenter, N. C., Horner, M. T., & Bernerth, J. B. (2012). Fairness at the collective level: A meta-analytic examination of the consequences and boundary conditions of organizational justice climate. *Journal of Applied Psychology*, *97*(4), 776-791.

Wilkesmann, U., & Lauer, S. (2015). What affects the teaching style of German professors? Evidence from two nationwide surveys. *Zeitschrift für Erziehungswissenschaft*, *18*(4), 713-736.

6 Study 3: Student attitudes toward fair assessment feedback: Results from a factorial survey

Burger, R., & Groß, M. (2016). *Student attitudes toward fair assessment feedback: Results from a factorial survey*. Manuscript submitted for publication.

Abstract This study investigates what constitutes fair feedback from a student perspective. Data was collected from 1,129 students in 46 departments of a German university in a factorial survey experiment. First, we investigate how fairness ratings are produced by the interplay of various attributes of a feedback situation. Next, we consider how the need for different types of feedback is defined by the grade a student has received. Finally, we examine how feedback expectations differ between disciplines. Regression results show that extent of feedback on how to improve has the largest contribution to fairness ratings, followed by feedback comprehensibility. In addition, the extent of feedback that students expect increases as grades decrease. As for disciplinary differences, the increase in demand for information regarding the grading criteria that follows poor grades is particularly large in soft pure disciplines. Results underline the benefits of using a factorial survey to study student attitudes.

6.1 Introduction

The importance of assessment feedback for student learning has received considerable attention in the higher education literature. Feedback serves as a guideline for students to identify their strengths and weaknesses. It offers them orientation on how to regulate their learning efforts in order to make progress (Black and Wiliam 1998; Hattie and Timperley 2007; Nicol and Macfarlane-Dick 2006). Feedback is therefore seen as a fundamental element of the assessment process (Evans 2013).

   While the possible benefits of assessment feedback are clearly evident, research on the subject has highlighted a range of problems that keep feedback from living up to its full potential. Of particular interest is the frequent observation that a large number of students expresses dissatisfaction with the feedback they receive (Ferguson 2011; Hounsell et al. 2008). On the other side of things, instructors often lament their students' reluctance to make use of feedback (Carless 2006; Mulliner and Tucker 2015). Recent studies have linked students' willingness to act on feedback to their emotional reactions to it (Pitt and Norton 2016; Robinson et al. 2013). Small and Attree (2015) point out that the chances that feedback is rejected are higher if it is contrary to students' expectations and past experiences. Accordingly, there is little to gain from providing feedback that seems appropriate from the instructor's perspective, but fails to address the students' needs (Orsmond and Merry 2011).

   The present study aims to enhance our understanding of the student perspective by asking what constitutes *fair* feedback from their point of view. In organizational justice research, the importance of a fair information policy has received attention under the heading *informational justice* (Greenberg 1993; see also Colquitt 2001). Informational justice "focuses on the explanations provided to people that convey information about why procedures were used in a certain way or why outcomes were distributed in a certain fashion" (Colquitt et al. 2001, p. 427). Thus far, justice perceptions in education have rarely been at the center of attention in empirical studies. This is particularly true with regard to student perceptions of the fairness of the procedures used in education (e.g. grading of assessments) and the perceived fairness of the explanations related to these procedures (e.g. assessment feedback) (Sabbagh and Resh 2016). At present, only one study is available that explicitly addresses informational justice in education (Kazemi

2016). There, Kazemi found a positive relationship between perceptions of informational justice and student achievement which is mediated via an increased motivation to study.

Greenberg (1993) points out that accurate and transparent explanations ensure that a distributive procedure is considered to be fair. This also applies in a higher education setting: If students are provided with insufficient information, they can feel that grading is biased and inconsistent (Carless 2006). It follows from this that feedback that is seen as unfair provokes a negative emotional response, potentially affecting students' inclination to engage with feedback.

This paper uses the concept of informational justice to describe the degree to which feedback is in accordance with student expectations and satisfies the informational needs that transpire from a given assessment situation. In order to reduce feelings of injustice, it is essential to achieve an empirically grounded understanding of how different factors are involved in determining how well feedback fulfills the above conditions from a student perspective. We approach this subject in three steps:

1. The first part is concerned with how perceptions of informational justice are produced by the interplay of various attributes of a feedback situation. The main interest here lies in the relative importance students assign to different feedback qualities.

2. Next, we widen the scope by investigating the degree to which expectations regarding the extent of feedback information are determined by the affordances of a particular assessment situation. To be precise, we seek to clarify how the need for different types of feedback information is defined by the grade a student has received.

3. Finally, we examine how student expectations are shaped by characteristics of curriculum and assessment process in different disciplinary environments.

This study uses a *factorial survey* to measure attitudes toward fair assessment feedback of a large sample of German university students ($n = 1,129$). This type of multifactorial experimental design was popularized by Peter H. Rossi (1979) and has proved to be particularly suited for the analysis of fairness-related judgment principles (for an excellent introduction, see Auspurg and Hinz 2015). While factorial surveys have been previously used on student samples (for a review, see Wallander 2009), there are currently no studies

available that apply this method to research questions specific to issues in higher education. Therefore, the secondary objective of this article is to familiarize higher education researchers with the advantages of this approach.

In a factorial survey, participants are presented several short, hypothetical scenarios, called *vignettes*. A vignette describes multiple attributes (*dimensions*) of a situation whose values (*levels*) are systematically varied between vignettes. Each respondent is assigned a number of vignettes whose fairness they have to rate. In the present case, vignettes describe situations in which a student has received feedback from an instructor following a graded assignment task. The vignette dimensions portray characteristics of the feedback as well as of the context in which it is given. By using regression analysis, we can retrace how vignette ratings are produced by the complex interplay of the vignette dimensions. This provides us with a refined and comprehensive picture of the multidimensional structure of expectations that generates perceptions of informational justice.

6.2   What makes feedback fair?

The following paragraphs introduce the various feedback characteristics that are expected to be involved in determining student perceptions of informational justice. First, we take a look at the importance ascribed to the extent of information provided by the instructor. Next, we consider the role of additional feedback qualities such as timeliness and comprehensibility.

With regard to feedback *extent*, this paper distinguishes two broad categories to which the content of feedback information can be assigned: descriptive feedback and constructive feedback. It is assumed that the extent with which feedback covers both types of content contributes to perceptions of informational justice. We use the term *descriptive feedback* to describe feedback that is concerned with explanations of grading criteria and of how these criteria were applied. Carless (2006) reports that students frequently have problems understanding grading criteria. Descriptive feedback helps to make the assessment process transparent and thus bestows legitimacy upon grading decisions (Lizzio and Wilson 2008). If feedback does not address these matters in a way that satisfies students' informational needs, it is expected that feelings of unjust treatment will be provoked.

Constructive feedback refers to information that provides input on where students should direct their efforts to improve learning outcomes. Since this type of feedback targets future progress, it has also been referred to as *feed-forward* (Evans 2013). The grade itself might serve a constructive function as well, but only to a point (Taras 2002). A

student who received a poor grade should interpret this information as a prompt to do better. But this prompt is of little use if they are not told *where* to do better, and, maybe more importantly, *how* to do this (Whitington et al. 2004). Hence, students are expected to experience injustice if feedback fails to adequately address these issues.

Apart from feedback extent, we assume that perceptions of informational justice are also influenced by a number of additional feedback characteristics. Comprehensibility and timeliness are commonly included as items in instruments measuring informational justice (e.g. Colquitt 2001, at present the most widely used scale). There is substantial evidence of the relevance of these dimensions in higher education contexts. Previous research has identified problems with *understanding* feedback as a major source of student dissatisfaction (Higgins et al. 2001; Williams 2005). Even very detailed feedback is less useful if recipients are unable to decipher its meaning. Feedback should also arrive in a *timely* manner so that students can actually benefit from it (Bayerlein 2014; Gibbs and Simpson 2004). If too much time passes between assessment and feedback, students could gradually lose interest in the subject (Denton et al. 2008). Lastly, students appear to prefer *verbal* feedback over written communication (Mulliner and Tucker 2015; O'Donovan et al. 2015; Orsmond et al. 2005). Feedback that is given in face-to-face interactions places students in an active dialogue with their instructors, which makes it easier to target specific issues that the students are interested in. Therefore, perceptions of informational justice are expected to be lower if the feedback in the vignette is described as hard to understand, comes with some delay, and is in written form instead of face-to-face.

Taken together, the implications of each individual dimension for perceptions of informational justice are rather straightforward. However, the *relative* contribution of each dimension to the *global* emotional response is less easy to anticipate. Is getting extensive constructive feedback valued more highly than descriptive feedback? How big is the impact of feedback extent compared to other feedback qualities? Thus far, research on these questions has been limited. In a notable exception, Winstone et al. (2015) used a budgeting methodology to study the relative importance of specific feedback topics. This approach shows some similarities to a factorial survey in that the importance of various dimensions is assessed in relation to each other. However, the cognitive challenge of weighing several dimensions against one another is subtler and more implicit in a factorial survey. A vignette stimulates a single rating that captures the multidimensionality of a situation instead of asking for exact, quantifiable values for individual dimensions

(Liebig et al. 2015). This enables a more realistic emulation of the thought process that determines perceptions of real-life assessment feedback.

6.2.1   Situational adjustment of feedback expectations

The central assumption of this study is that feedback is seen as fair if it meets student expectations. We further assume that these expectations are not constant across feedback situations. Rather, they reflect informational *needs* that are created by and specific to a particular assignment. While this has not yet been investigated in a systematic fashion, previous research does give some supporting evidence for this line of reasoning. Crisp (2007) argues that one reason why students do not engage with feedback is that they are less interested in feedback if their grades are good. This type of moderation needs to be accounted for when asking about attitudes toward fair feedback: Do students expect more detailed explanations of the reasoning behind a grade if the grade was bad? How does the grade relate to the need for information on how to improve? Indeed, Liebig et al. (2015) posit that individuals need contextual information to be able to assess the fairness of a specific treatment. With that in mind, information on the *grade* of the assignment for which feedback is given was included as a vignette dimension.

The grade for an assessment provides information on the quality of student effort in a very condensed form. Descriptive feedback disentangles this information by explaining how various criteria were applied in the grading process to arrive at the result. If this subject is not covered by the feedback, the assessment process appears less transparent, which should lower perceptions of informational justice. It is now suggested that the degree to which justice perceptions suffer depends on the *need* to know how the grade is justified. Interest in the reasoning behind a grading decision should be lower if the grade was good. But the lower the grade, the more compelling the desire to understand the result, and thus the demand for descriptive feedback (Ferguson 2011). If this need is not recognized by the instructor, feelings of injustice are expected to increase.

It is expected that the grade also influences the need for constructive feedback, albeit for different reasons. One of the prime motivators for giving constructive feedback is to supply students with information they can use to improve their abilities and to make progress (Nicol and Macfarlane-Dick 2006). It is plausible that the demand for such information should depend on the grade a student received: The worse the grade, the more room for improvement there is. Not only that, the worse the grade, the more pressing the actual *need* to improve. Therefore, it is expected that students who are awarded a lower

grade exhibit a higher demand for receiving extensive constructive feedback. If instructors fail to give feedback of this type, a greater sense of injustice is provoked than if the grade was good.

6.2.2 Disciplinary differences

The previous paragraphs described how expectations of descriptive and constructive feedback are dynamically adjusted to the context in which feedback is given. Another layer can be added to this argument by proposing that the relationship between grade and need is also a reflection of previous experiences with assessment and feedback. Given that curriculum and assessment procedures are subject to considerable variation between academic disciplines (Becher 1989; Neumann 2001; Ramsden 1979), it is likely that feedback-related experiences (and thus expectations) vary as well. There is a noticeable lack of research in this area. As Evans (2013) points out in her comprehensive literature review, the majority of research on assessment feedback was conducted in a small, select number of disciplinary contexts. In cases where multiple disciplines were investigated, potential differences are rarely discussed (for an exception, see Fernández-Toro et al. 2013).

To investigate the role of the disciplinary context, the present study employs Becher's (1989) classification scheme, which in turn is based on work by Biglan (1973a, 1973b) and Kolb (1981). This framework distinguishes hard pure, hard applied, soft pure, and soft applied disciplines. According to Neumann et al. (2002), the type of knowledge that is the foundation of hard pure (e.g. physics, biology) and hard applied (e.g. engineering, medicine) disciplines favors a more standardized curriculum based around teaching established facts. This is underlined by a strong reliance on assessment via "objective" tests such as multiple-choice exams. The curriculum in soft pure (e.g. history, philosophy) and soft applied (e.g. business, law) disciplines is less concerned with teaching facts. Rather, students are expected to develop skills like creative thinking and the ability to express their own ideas. These abilities are harder to assess via standardized tests, which is why there is a stronger inclination toward methods such as assignment essays and open-question exams.

These divergent properties have some important implications for how decreasing grades should increase the demand for descriptive feedback. We propose that there are different motivating factors behind this moderation that reflect the fact that descriptive feedback serves different functions in different contexts. First, descriptive feedback is

instructive in that it caters to a general interest to know where wrong answers were given. Second, descriptive feedback is *legitimizing* in that it alleviates concerns regarding the validity of the grading process: Was a poor grade due to a lack of effort, or was it because the instructor does not share the student's view on the subject matter, has a dislike of the student, etc.?

The need for both functions is expected to be inversely related to the grade. But while instruction is relevant in all disciplines, legitimization is particularly necessary in environments where validity-related questions are likely to develop. In *hard disciplines*, the instructive function prevails. While students in these fields are assumed to have a higher demand for instruction on where they gave wrong answers following a sub-par assignment, they probably do not doubt that their answers were, in fact, wrong. In *soft disciplines*, grading criteria are more ambiguous, as it is harder to specify what constitutes a right or wrong answer. Since grading such work leaves more room for subjectivity, lower grades are more likely to cast doubts on the validity of the outcome. Thus, descriptive feedback is not only needed for instruction, but also for legitimization. Because of that, the increase in demand for descriptive feedback following a less-than-good grade is hypothesized to be more pronounced in soft disciplines than in hard disciplines. It follows from this that the theoretical focus is on the hard-soft dichotomy. Nevertheless, the analyses will distinguish all four disciplinary groups in order to detect potential differences along the pure-applied axis.

In order to contrast the respective status of descriptive and constructive feedback across disciplines, we also check for the presence of disciplinary differences in the moderation of constructive feedback. While the divergent properties of curriculum and assessment in hard and soft fields have clear implications for how demand for descriptive feedback is determined by the grade, it is harder to formulate a corresponding hypothesis for constructive feedback. The previous paragraphs suggest that descriptive feedback serves extended functions in soft disciplines that are less critical in hard disciplines. However, it can be argued that the greater need for constructive feedback caused by lower grades is not tied to a specific assessment format or curriculum, and thus of a more general nature. Therefore, disciplinary differences are expected to be less pronounced for the moderation of constructive feedback by grade.

6.3 Data and method

6.3.1 Participants

The factorial survey was set up as an online questionnaire. In order to ensure the adequacy of the research design, a pilot study was carried out among a sample of $n = 196$ German university students recruited from a commercial sampling pool. Afterwards, the main study was conducted among $n = 1,129$ students from $n = 46$ departments of a large German university. Each respondent rated 10 vignettes, so that the number of cases available for analysis totals to 11,290. Participants were recruited via the university mailing list. 66.5% of respondents were female. Mean age was 23.7 years (SD = 4.5). On average, participants had been enrolled for 5.3 semesters at the time of the survey (SD = 3.4). 25.8% of participants were enrolled in hard pure (e.g. biology, mathematics), 23.2% in hard applied (e.g. medicine, computer science), 32% in soft pure (e.g. history, philosophy), and 19% in soft applied disciplines (e.g. law, business).

6.3.2 Vignettes

The vignette module was prefaced by an introduction that sets the framing and explains the procedure. Participants were asked to carefully read the vignette texts and to rate the fairness of the feedback. Based on recommendations in Auspurg and Hinz (2015), an 11-point rating scale was used (1: not fair at all – 11: very fair). Vignettes appear as full, syntactically correct sentences. This supports the goal of presenting realistic situations that closely resemble the students' own experiences with assessment feedback. Each vignette has two distinct components. The first component describes the attributes of the assignment for which feedback was given. The second part details properties of the feedback given by the instructor. A list of all vignette dimensions and levels appears in Table 1. After finishing the vignette module, participants were asked a number of questions regarding their study-related experiences and attitudes as well as their socio-demographic background.

Example vignette

> [Translated from original German] A student has received an *average* grade for an *essay*. They spent *several weeks* preparing for this assignment. The feedback provided *extensive* explanations regarding the grading criteria. It contained *no* information that addressed potential for improvement. The explanations came *after a couple of days* in *personal conversation* and were *hard* to understand.

> How do you rate the fairness of the lecturer's information policy in this situation?

Table 6-1 Vignette dimensions and levels.

Dimension	Levels
Grade	good / average / poor
Assessment method	exam / essay
Student effort	a couple of days / several weeks
Descriptive feedback	extensive / short / no
Constructive feedback	extensive / short / no
Timeliness	a couple of days / several weeks
Feedback mode	personal conversation / written form
Comprehensibility	easy / hard

6.3.3 Experimental design

The vignettes contain eight dimensions, five of which have two levels and three have three levels. This gives us $2^5 \times 3^3 = 864$ total possible configurations of vignette dimensions (*vignette universe*). A fractional design was used where a sample of N = 240 was drawn from the vignette universe using the %Mktex macro (Kuhfeld 2010) for the software package SAS. This type of sampling produces a D-efficient design that maximizes orthogonality of vignette dimensions and balance of levels. This increases the precision with which regression parameters can be estimated over what is possible using random sampling (Auspurg and Hinz 2015). A D-efficiency of D = 99.17 was reached, slightly less than the possible maximum of 100. Each vignette was blocked to one of 24 decks, yielding 10 vignettes per deck. The blocking step used the same efficiency algorithm as the sampling procedure. Decks were randomly assigned to respondents at survey runtime. The sequence of vignettes was randomized for each respondent.

6.3.4   Procedure

The hierarchic nature of the data (vignette ratings nested in respondents) suggests the use of linear mixed effect regression models that include a respondent-specific random intercept. The multilevel-approach ensures that intra-respondent correlation of vignette ratings is factored into the estimation procedure. Failing to account for the hierarchical structure of the data would lead to wrong estimates of the standard errors (Raudenbush and Bryk 2002; Snijders and Bosker 2012). Regression models were estimated using version 1.1-11 of the lme4 package (Bates et al. 2015) in R 3.2.4.

The vignette ratings are used as the dependent variable in multivariate regression analyses, independent variables are given by the vignette dimensions. Systematic differences in ratings between vignettes with different levels of a dimension allow us to identify the impact of that dimension (Auspurg and Hinz 2015). Models also include categorical variables for the vignette deck as well as a numerical variable indicating the sequential position in which a vignette appeared in the vignette module. Additional control variables on the respondent level are gender, year of study, student's own experiences with feedback, and satisfaction with achievement.

6.4   Results

6.4.1   Relative importance of feedback dimensions

What actually constitutes fair feedback from a student perspective? Fig. 6-1 visualizes the results of the first multivariate regression model. Here, we see the respective contributions of each feedback dimension to the overall vignette ratings. While each dimension exerts significant influence on ratings of informational justice ($p < .001$), it is clear that the *relative* size of the effects varies substantially. Let us first consider the effects of the extent of information on both types of feedback content. If descriptive feedback in a vignette was short instead of extensive, participants gave ratings that were, on average, .941 points lower ($t = -22.144$, $p < .001$), all other variables held constant. If no explanations regarding grading criteria were given, ratings were lower by 1.959 on the 11-point scale ($t = -46.678$, $p < .001$). The negative impact on fairness ratings is even more pronounced for constructive feedback. The average difference in ratings of vignettes with short versus vignettes with extensive constructive feedback is estimated to be 1.257 ($t = -29.984$, $p < .001$). If information on how to improve was not part of the feedback at all, ratings were lower by 2.581 points ($t = -60.573$, $p < .001$). Thus, insufficient explanations in both

thematic categories lead to a substantial increase in the likelihood that students dismiss the feedback on grounds of being unfair.
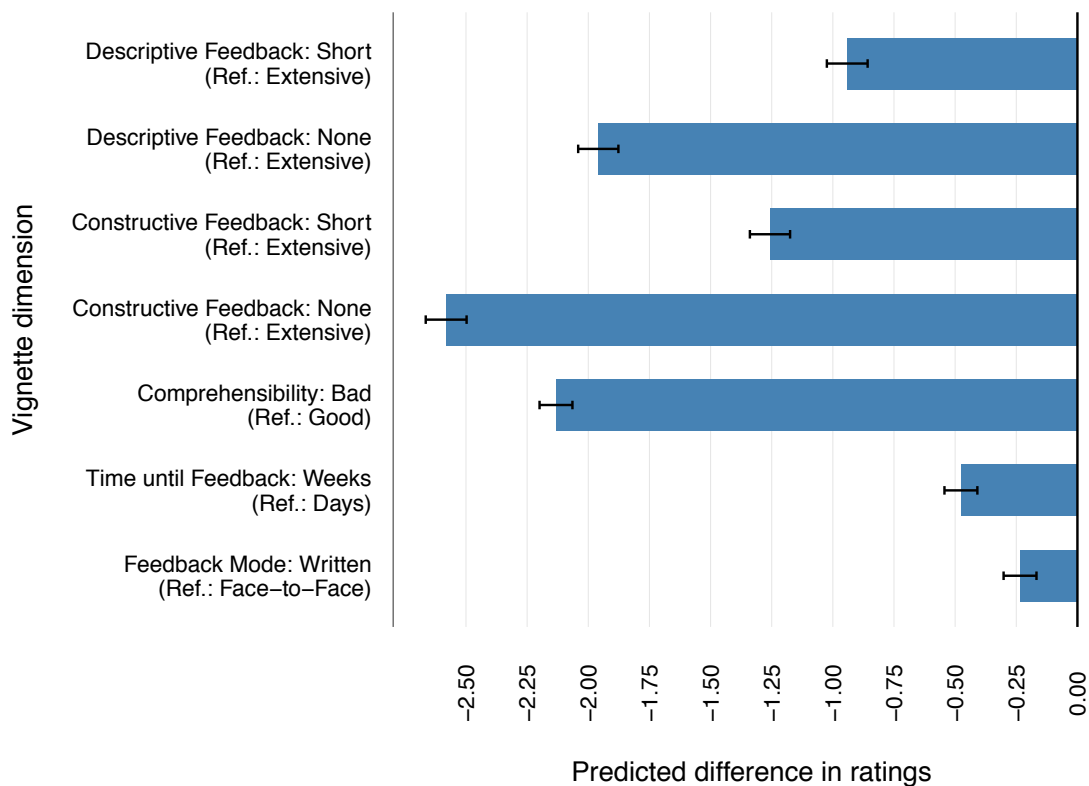

Fig. 6-1 Effects of vignette dimensions on ratings of informational justice. 95% confidence bands.

Results for the additional feedback qualities also support the hypotheses. If feedback was described as hard to understand, students gave ratings that were lower by 2.132 points ($t = -62.161$, $p < .001$). Ratings decreased by .476 if the vignette said that feedback was delivered after several weeks instead of a couple of days ($t = -13.884$, $p < .001$). Lastly, verbal feedback was preferred over written feedback, with the latter causing vignette ratings to drop by .235 points ($t = -6.844$, $p < .001$). In light of previous research, it is not surprising that students prefer feedback that is timely and delivered face-to-face. Nevertheless, it is certainly interesting that the contribution of these dimensions to the overall perception of informational justice is almost negligible when compared to the effects of feedback extent and comprehensibility.

6.4.2   Situational adjustment of feedback expectations

Descriptive Feedback

The next step of the analyses is concerned with how feedback-related expectations are dynamically adjusted to the affordances of the feedback situation. An interaction term between extent of descriptive feedback and grade is added to the multivariate regression model. It was hypothesized that expectations regarding the extent of descriptive feedback should increase with decreasing grades. The Results shown in Fig. 6-2 lend partial support to this hypothesis. We see that both average and poor grades tend to elicit lower ratings of informational justice than good grades. However, for *average* grades, the magnitude of the decrease in ratings is not significantly different from good grades, as is evidenced by the lines running almost parallel. For *poor* grades, we see significant moderation. Compared to vignettes in which the student had received a good grade, the decrease in ratings of informational justice was larger by .348 ($t = -3.176$, $p < .01$) if descriptive feedback was short and by .229 ($t = -2.095$, $p < .05$) if no explanations of the grading procedure were provided, all other variables held constant.
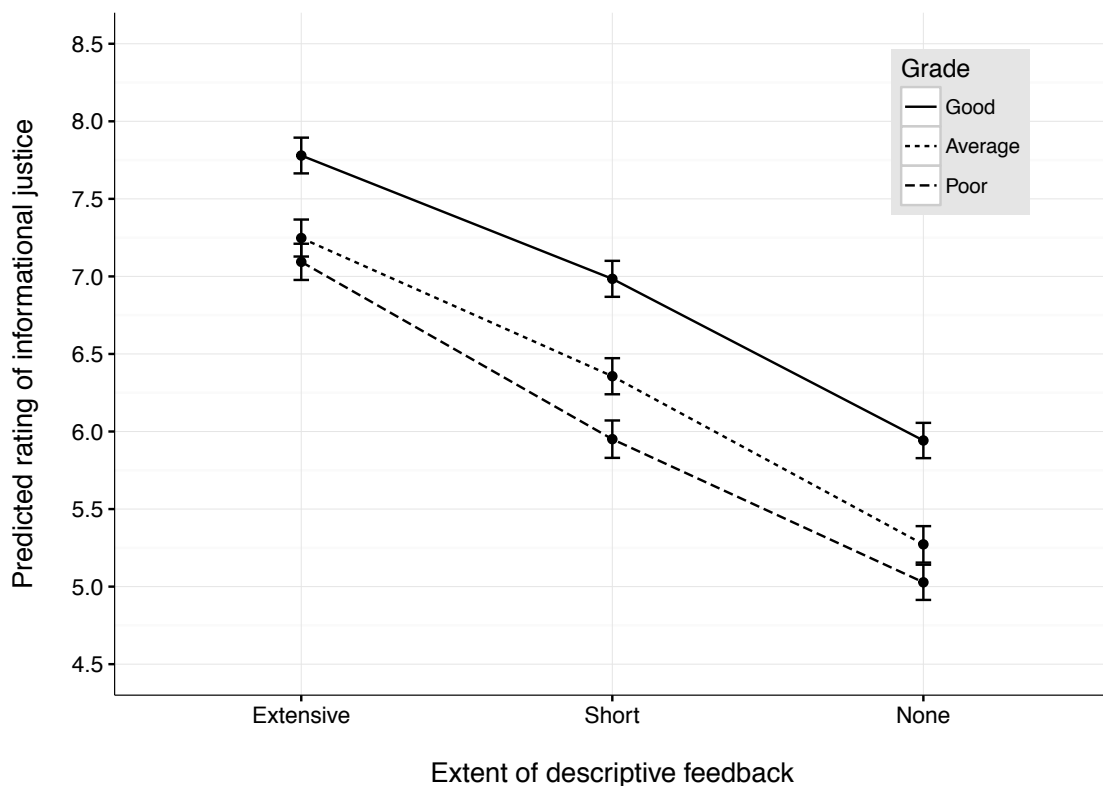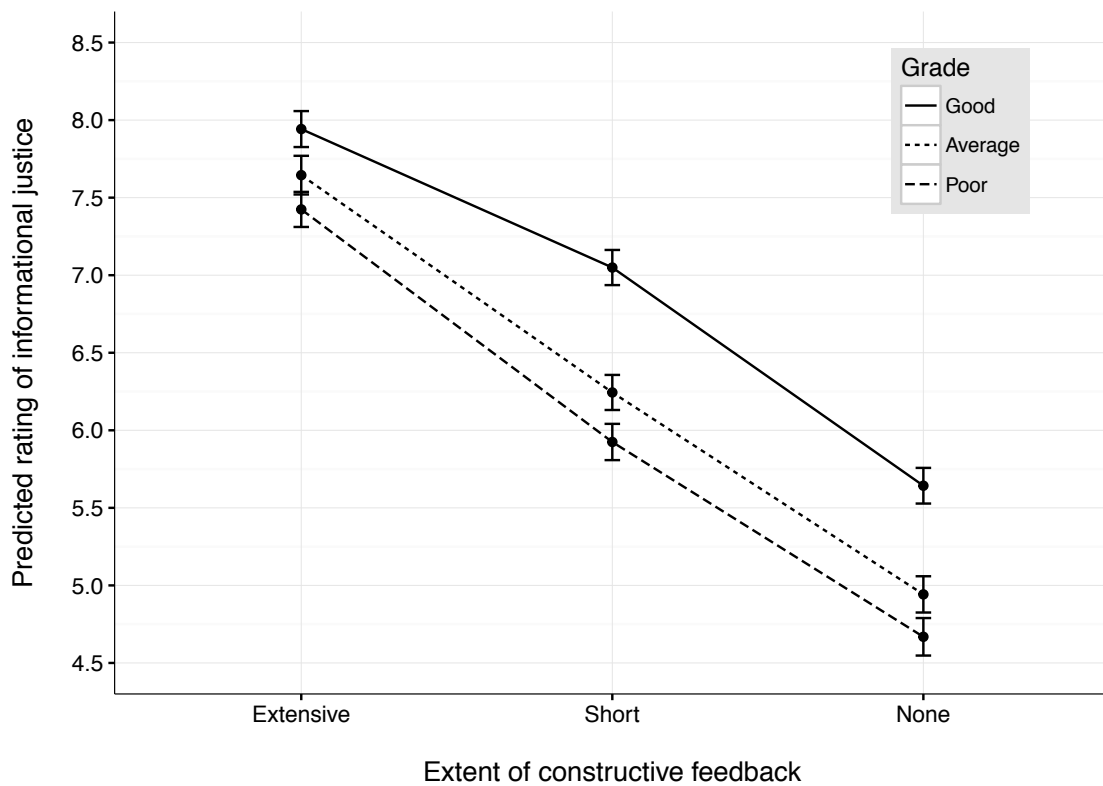

Fig. 6-2 Effect of extent of descriptive feedback on ratings of informational justice for different grades. 95% confidence bands.

Constructive Feedback

The next multivariate regression model tests for moderation of constructive feedback by grade (Fig. 6-3). Compared to the interaction between grade and descriptive feedback, interaction effects are significant for both average and poor grades ($p < .001$). As long as feedback is extensive, differences by grade are rather small. But short constructive feedback is only seen as somewhat acceptable if it is given as a follow up for a good grade. If short feedback was provided for an *average* grade, the decrease in ratings was greater by .508 ($t = -4.537$, $p < .001$) than for good grades; in case of no constructive feedback the difference was .403 ($t = -3.567$, $p < .001$). In vignettes where the grade was *poor* and there was short constructive feedback, the decrease in ratings was greater by .607 ($t = -5.639$, $p < .001$) relative to vignettes where the grade was good; in case of no constructive feedback the difference was .456 ($t = -4.260$, $p < .001$). This supports the argument that lower grades not only leave more room for improvement. They also create a distinct need to improve, as is evidenced by the greater demand for feedback that caters to this need. If instructors fail to account for this, a greater sense of injustice is provoked.

Fig. 6-3 Effect of extent of constructive feedback on ratings of informational justice for different grades. 95% confidence bands.

6.4.3   Disciplinary differences

For the final part of our analyses, the full sample used in the previous regression models was divided into four subsamples. Each participant was assigned to one of four disciplinary groups (hard pure, hard applied, soft pure, soft applied). Regression models for situational adjustment of feedback expectations were re-estimated separately for each group.

Descriptive Feedback

First, we look at interactions with extent of descriptive feedback (Table 6-2). In the full-sample model, the interaction between grade and descriptive feedback was significant for poor grades, but not for average grades. It was expected that this interaction should be stronger in soft disciplines than in their hard counterparts because descriptive feedback in soft fields is not only instructive, but also serves to legitimize the grading decision. The results give some evidence in favor of this hypothesis. The moderation in soft pure fields (e.g. English, art history) is not only stronger than in other disciplines, it is actually exclusively observed among students in this subsample. If a vignette described a situation in which the student had received short descriptive feedback following an *average* grade, the decrease in ratings was larger by .477 ($t = -2.446$, $p < .05$) relative to good grades for participants from soft pure disciplines; in case of no descriptive feedback the difference was .577 ($t = -2.946$, $p < .01$). For *poor* grades, the decrease in ratings was larger by .695 ($t = -3.513$, $p < .001$) for short and by .590 ($t = -3.044$, $p < .01$) for no descriptive feedback. Since these interactions are not present in the other disciplines, we conclude that the need for instruction via descriptive feedback is not tied to the grade, while the need for legitimization is. The unexpected findings in soft applied fields will be put under closer scrutiny in the discussion.

Table 6-2 Moderation of effect of descriptive feedback on ratings of informational justice by grade. Overall model and disciplinary subsamples. Standard errors in parentheses.

	Overall b / SE	Hard pure b / SE	Hard appl. b / SE	Soft pure b / SE	Soft appl. b / SE
Grade (*Ref.: Good*)					
Average	-.532***	-.635***	-.527***	-.211	-.900***
	(.076)	(.147)	(.156)	(.138)	(.167)
Poor	-.686***	-.700***	-.834***	-.381**	-.934***
	(.076)	(.155)	(.154)	(.135)	(.166)
Descriptive feedback (*Ref.: Extensive*)					
Short	-.795***	-.913***	-.959***	-.456***	-.978***
	(.075)	(.147)	(.154)	(.135)	(.163)
None	-1.837***	-1.953***	-2.001***	-1.560***	-1.964***
	(.075)	(.153)	(.152)	(.133)	(.164)
Interactions					
Short * Average	-.096	.139	-.096	-.477*	.170
	(.108)	(.216)	(.223)	(.195)	(.241)
Short * Poor	-.348**	-.238	-.127	-.695***	-.263
	(.110)	(.220)	(.224)	(.198)	(.240)
None * Average	-.138	.063	-.015	-.577**	.128
	(.108)	(.216)	(.221)	(.196)	(.238)
None * Poor	-.229*	-.288	.028	-.590**	.109
	(.109)	(.224)	(.224)	(.194)	(.240)
n respondents	1129	291	262	361	215
n vignettes	11290	2910	2620	3610	2150

Additional controls (vignette level): assessment type; student effort; constructive feedback; timeliness; feedback mode; comprehensibility; vignette order.

Additional controls (respondent level): gender; year of study; experiences with feedback; satisfaction with achievement; vignette deck

***$p < .001$; **$p < .01$; *$p < .05$

Constructive Feedback

How do disciplinary differences for constructive feedback compare to descriptive feedback? The argument was that the increase in need for constructive feedback following lower grades is not particular to a specific type of curriculum. Therefore, disciplinary differences should be less pronounced than for descriptive feedback. Results of the regression models containing interactions between extent of constructive feedback and grade for each disciplinary group are presented in Table 6-3.

Table 6-3 Moderation of effect of constructive feedback on ratings of informational justice by grade. Overall model and disciplinary subsamples. Standard errors in parentheses.

	Overall *b / SE*	Hard pure *b / SE*	Hard appl. *b / SE*	Soft pure *b / SE*	Soft appl. *b / SE*
Grade (*Ref.: Good*)					
Average	-.297***	-.234	-.258	-.197	-.595***
	(.080)	(.161)	(.165)	(.143)	(.176)
Poor	-.518***	-.494***	-.380*	-.547***	-.632***
	(.074)	(.147)	(.151)	(.135)	(.161)
Constructive feedback (*Ref.: Extensive*)					
Short	-.893***	-.875***	-.776***	-.866***	-1.076***
	(.075)	(.150)	(.154)	(.133)	(.165)
None	-2.299***	-2.270***	-2.284***	-2.327***	-2.296***
	(.075)	(.149)	(.153)	(.137)	(.159)
Interactions					
Short * Average	-.508***	-.570*	-.588*	-.636**	-.128
	(.112)	(.224)	(.230)	(.200)	(.249)
Short * Poor	-.607***	-.650**	-.926***	-.476*	-.385
	(.108)	(.216)	(.221)	(.193)	(.237)
None * Average	-.403***	-.411	-.291	-.430*	-.484*
	(.113)	(.225)	(.232)	(.205)	(.246)
None * Poor	-.456***	-.503*	-.520*	-.273	-.671**
	(.107)	(.215)	(.218)	(.193)	(.236)
n respondents	1129	291	262	361	215
n vignettes	11290	2910	2620	3610	2150

Additional controls (vignette level): assessment type; student effort; descriptive feedback; timeliness; feedback mode; comprehensibility; vignette order.

Additional controls (respondent level): gender; year of study; experiences with feedback; satisfaction with achievement; vignette deck

***$p < .001$; **$p < .01$; *$p < .05$

We see that the effect of constructive feedback is moderated by the grade in all disciplines. If constructive feedback is not extensive, ratings of informational justice suffer. This decrease gets stronger as grades get lower. With the exception of soft applied disciplines, the size of the interaction effects is comparable to what we saw in the full-sample model. Some interactions fail to reach statistical significance. This is true for interactions between no constructive feedback and average grade in hard pure ($t = -1.824$, $p = .069$) and hard applied fields ($t = -1.253$, $p = .211$) and the interaction between no constructive feedback and poor grade in soft pure fields ($t = -1.417$, $p = .157$). In soft applied fields, interactions are not significant if short constructive feedback was given ($t = -.513$, $p = .608$ for average grades; $t = -1.629$, $p = .104$ for poor grades), but they are for no constructive feedback ($t = -1.969$, $p < .05$ for average grades; $t = -2.841$, $p < .01$ for poor grades).

Even though it is hard to establish a clear pattern, we find evidence that the need to receive information that aims to improve learning outcomes depends on the grade in all disciplines. This is in contrast to the findings for descriptive feedback, where the moderation by grade seems to be primarily motivated by a need for legitimization, which is present only in soft pure fields. Therefore, we conclude that both types of feedback content not only cater to different needs, but that these needs also take a different shape depending on the disciplinary context.

6.5 Discussion

The purpose of this study was to enhance our understanding of the mechanisms that determine whether or not assessment feedback appears fair from a student perspective. This was motivated by the idea that students should be more likely to accept feedback that is in accordance with their informational needs and does not provoke a negative emotional reaction. The present study contributes to existing research on assessment feedback in three areas. First, we showed the relative contribution of various feedback characteristics to ratings of informational justice. Next, we demonstrated how the need for feedback depends on the grade. Finally, we considered how student expectations vary between academic disciplines. Taken together, the results of the factorial survey offer valuable insights into the complex interplay of factors that underlies student perceptions of informational justice.

With regard to the weight assigned to individual feedback dimensions, it is clear that students feel entitled to feedback that is extensive as well as comprehensible. The extent of both descriptive and constructive feedback exerts considerable influence on fairness ratings, with more weight placed on information on how to improve. The preference for constructive over descriptive feedback found in the factorial survey is in support of results from Winstone et al. (2015). The importance of timeliness and feedback mode pales in comparison to the other dimensions. To better illustrate this: The intercept in the first regression model was estimated at 10.890. This means that for feedback that is aligned with student expectations in all dimensions, ratings are very close to the upper end of the 11-point scale. Starting from this number, ratings of feedback that comes in written form as well as with some delay is predicted to be around 10.179, all other variables held constant. While this is certainly lower than the baseline, the overall sentiment is still very much on the fair side of things. But if, for example, feedback is said to omit information on how to improve and is at the same time hard to understand, the emotional

response is decidedly more negative. These results provide a strong case for employing a factorial survey to investigate attitudes toward assessment feedback. A conventional, item-based survey could not have captured how *global* perceptions of a situation are produced by the interplay of multiple dimensions. Timely and face-to-face feedback are important, but they are no substitute for content quality. It would be ill-advised to rush out feedback if this means sacrificing content. This echoes findings by Ferguson (2011). Similarly, the preference for face-to-face feedback might cast doubts on the viability of delivering feedback via electronic means. But in the grand scheme of things, negative sentiments towards written feedback are easily compensated by quality content.

Moving on to the next step, we found substantial evidence that students adjust their informational demands to the needs created by a particular assignment. The lower the grade for an assignment, the greater the need to know the reasoning behind the grade and to receive directions for future improvement. This contextual dependence is more prominent for constructive feedback than it is for descriptive feedback. Again, this part of the research agenda benefitted from the opportunities offered by the factorial survey method. Contextual information is easily accommodated in the vignettes. This places the vignette situation closer to the students' own experiences – it is hard to imagine that feedback is evaluated free of context in real-life situations. Interestingly, a lack of both descriptive and constructive feedback puts a major penalty on fairness perceptions even if the grade was good. This adds a new perspective on previous findings that suggest that students do not care about feedback if they are happy with their grade (Crisp 2007). Even though good grades create less demand for feedback, the respective extent of either feedback category remains a major contributor to the global justice judgment. While instructors should take special care of the amount of feedback they give for lower grades, they should not refrain from supplying their better students with at least short explanations.

As for the disciplinary differences, results highlight the necessity of accounting for the disciplinary context when investigating attitudes toward assessment feedback. Even though the sentiment that feedback needs to address the grading criteria is present in all disciplines, the demand for information of this type is only tied to the grade in soft pure fields. This was explained by the greater need for legitimizing explanations for lower grades in contexts where assessment is based around discourse and argumentation instead of established facts. If only the full sample had been used, the status of descriptive feedback in soft pure fields would have been severely underestimated. Conversely, we would

have concluded that the need for descriptive feedback depends on the grade in all disciplines, when in fact this does not seem to be the case.

Regarding the latter point, findings in soft applied fields contradicted our hypothesis regarding the moderation of the effects of descriptive feedback by the grade. However, upon closer inspection of the departments in the sample, it appears that even though students from soft applied fields *should* demand more descriptive feedback, the actual situation in their departments might prevent the emergence of such expectations. As the theory suggests, assessment in both soft pure and applied fields in the sample relies more strongly on essays than it does in hard disciplines. Taken by itself, this is in support of our argument regarding the need for legitimization. However, when compared to soft pure fields, the student-faculty-ratio is much more unfavorable in soft applied fields. The proportion of classes that take place in lectures as opposed to smaller seminars is larger as well, resulting in fewer chances of direct interactions between students and lecturers (Cuseo 2007; Severiens et al. 2015). This places some structural constraints on the process of giving feedback. In light of that, the lack of a significant interaction is plausible: Because students in soft applied departments are less likely to experience instructors tailoring feedback to their needs, they do not consider the possibility that this should actually take place.

The generalizability of the findings reported here warrants some consideration. Data was collected from a large number of students from decidedly heterogeneous disciplinary backgrounds. As such, the results have robust support and can be considered to be representative for the underlying population. Nevertheless, the factorial survey was limited to students of a single German university. It is possible that the preferential hierarchy discovered here takes a different form in other contexts. Whether or not a situation is seen as fair depends on the individual frame of reference. This frame could be influenced by factors like organizational structure, size of the institution, and institutional policies. To get past this limitation, it is essential that further research investigates the proposed relationships among a more varied selection of student populations, especially from a cross-country perspective. Comparisons between institutions would not only tell us something about the external validity of our data, it would also allow us to identify how feedback-related expectations are shaped by specific features of higher education systems.

In conclusion, the results of this study provide a comprehensive picture of the mechanisms that shape student perceptions of assessment feedback. This can be used as

a starting point for devising feedback strategies that give more consideration to the students' needs and expectations, thus reducing feelings of injustice that are detrimental to the learning experience. Of course, it needs to be pointed out that the mere fact that students would consider a particular type of feedback to be fair does not necessitate that they will take action on the feedback. However, the chances that they will do so are arguably higher than if the feedback fails to meet their expectations. While substantial time and energy are required if instructors are to meet their students' needs in a manner that is not only superficial (Price et al. 2010), devoting more attention to issues of informational justice promises not only greater satisfaction among students, but also a lower risk that instructors' efforts are in vain.

6.6 References

Auspurg, K., & Hinz, T. (2015). *Factorial survey experiments*. Thousand Oaks, CA: Sage.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1-48.

Bayerlein, L. (2014). Students' feedback preferences: How do students react to timely and automatically generated assessment feedback?. *Assessment & Evaluation in Higher Education*, *39*(8), 916-931.

Becher, T. (1989). *Academic tribes and territories: Intellectual inquiry and the cultures of disciplines*. Milton Keynes: SRHE/Open University Press.

Biglan, A. (1973a). The characteristics of subject matter in different scientific areas. *Journal of Applied Psychology*, *57*(3), 195–203.

Biglan, A. (1973b). Relationships between subject matter characteristics and the structure and output of university departments. *Journal of Applied Psychology*, *57*(3), 204–213.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, *5*(1), 7-74.

Carless, D. (2006). Differing perceptions in the feedback process. *Studies in Higher Education*, *31*(2), 219-233.

Colquitt, J. A. (2001). On the dimensionality of organizational justice: A construct vali-
    dation of a measure. *Journal of Applied Psychology*, *86*(3), 386-400.

Colquitt, J. A., Conlon, D. E., Wesson, M. J., Porter, C. O., & 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.

Crisp, B. R. (2007). Is it worth the effort? How feedback influences students' subse-
    quent submission of assessable work. *Assessment & Evaluation in Higher Edu-
    cation*, *32*(5), 571-581.

Cuseo, J. (2007). The empirical case against large class size: Adverse effects on the
    teaching, learning, and retention of first-year students. *The Journal of Faculty
    Development*, *21*(1), 5-21.

Denton, P., Madden, J., Roberts, M., & Rowe, P. (2008). Students' response to tradi-
    tional and computer-assisted formative feedback: A comparative case
    study. *British Journal of Educational Technology*, *39*(3), 486-500.

Evans, C. (2013). Making sense of assessment feedback in higher education. *Review of
    Educational Research*, *83*(1), 70-120.

Ferguson, P. (2011). Student perceptions of quality feedback in teacher education. *As-
    sessment & Evaluation in Higher Education*, *36*(1), 51-62.

Fernández-Toro, M., Truman, M., & Walker, M. (2013). Are the principles of effective
    feedback transferable across disciplines? A comparative study of written assign-
    ment feedback in languages and technology. *Assessment & Evaluation in Higher
    Education*, *38*(7), 816-830.

Gibbs, G., & Simpson, C. (2004). Conditions under which assessment supports stu-
    dents' learning. *Journal of Teaching and Learning in Higher Education*, *1*(1), 1-
    30.

Greenberg, J. (1993). The social side of fairness: Interpersonal and informational clas-
    ses of organizational justice. In R. Cropanzano (Ed.), *Justice in the workplace:
    Approaching fairness in human resource management* (pp. 79-103). Hillsdale,
    NJ: Erlbaum.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Re-
    search*, *77*(1), 81-112.

Higgins, R., Hartley, P., & Skelton, A. (2001). Getting the message across: The problem of communicating assessment feedback. *Teaching in Higher Education*, *6*(2), 269-274.

Hounsell, D., McCune, V., Hounsell, J., & Litjens, J. (2008). The quality of guidance and feedback to students. *Higher Education Research & Development*, *27*(1), 55-67.

Kazemi, A. (2016). Examining the interplay of justice perceptions, motivation, and school achievement among secondary school students. *Social Justice Research*, *29*(1), 103-118.

Kolb, D. A. (1981). Learning styles and disciplinary differences. In A. Chickering (Ed.), *The modern American college* (pp. 232-255). San Francisco, CA: Jossey-Bass.

Kuhfeld, W. F. (2010). *Efficient experimental designs using computerized searches*. Sequim, WA: SAS Institute.

Liebig, S., Sauer, C., & Friedhoff, S. (2015). Using factorial surveys to study justice perceptions: Five methodological problems of attitudinal justice research. *Social Justice Research*, *28*(4), 415-434.

Lizzio, A., & Wilson, K. (2008). Feedback on assessment: Students' perceptions of quality and effectiveness. *Assessment & Evaluation in Higher Education*, *33*(3), 263-275.

Mulliner, E., & Tucker, M. (2015). Feedback on feedback practice: Perceptions of students and academics. *Assessment & Evaluation in Higher Education*. doi:10.1080/02602938.2015.1103365.

Neumann, R. (2001). Disciplinary differences and university teaching. *Studies in Higher Education*, *26*(2), 135-146.

Neumann, R., Parry, S., & Becher, T. (2002). Teaching and learning in their disciplinary contexts: A conceptual analysis. *Studies in Higher Education*, *27*(4), 405-417.

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, *31*(2), 199-218.

O'Donovan, B., Rust, C., & Price, M. (2015). A scholarly approach to solving the feedback dilemma in practice. *Assessment & Evaluation in Higher Education*. doi:10.1080/02602938.2015.1052774.

Orsmond, P., & Merry, S. (2011). Feedback alignment: Effective and ineffective links between tutors' and students' understanding of coursework feedback. *Assessment & Evaluation in Higher Education, 36*(2), 125-136.

Orsmond, P., Merry, S., & Reiling, K. (2005). Biology students' utilization of tutors' formative feedback: A qualitative interview study. *Assessment & Evaluation in Higher Education, 30*(4), 369-386.

Pitt, E., & Norton, L. (2016). "Now that's the feedback I want!" Students' reactions to feedback on graded work and what they do with it. *Assessment & Evaluation in Higher Education*. doi:10.1080/02602938.2016.1142500.

Price, M., Handley, K., Millar, J., & O'Donovan, B. (2010). Feedback: All that effort, but what is the effect?. *Assessment & Evaluation in Higher Education, 35*(3), 277-289.

Ramsden, P. (1979). Student learning and perceptions of the academic environment. *Higher Education, 8*(4), 411-427.

Raudenbush, S. W., & Bryk, A.S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). Los Angeles, CA: Sage.

Robinson, S., Pope, D., & Holyoak, L. (2013). Can we meet their expectations? Experiences and perceptions of feedback in first year undergraduate students. *Assessment & Evaluation in Higher Education, 38*(3), 260-272.

Rossi, P. H. (1979). Vignette analysis: Uncovering the normative structure of complex judgments. In R. K. Merton, J. S. Coleman, & P. H. Rossi (Eds.), *Qualitative and quantitative social research: Papers in honor of Paul F. Lazarsfeld* (pp. 176-186). New York, NY: Free Press.

Sabbagh, C., & Resh, N. (2016). Unfolding justice research in the realm of education. *Social Justice Research, 29*(1), 1-13.

Severiens, S., Meeuwisse, M., & Born, M. (2015). Student experience and academic success: Comparing a student-centred and a lecture-based course programme. *Higher Education, 70*(1), 1-17.

Small, F., & Attree, K. (2015). Undergraduate student responses to feedback: Expectations and experiences. *Studies in Higher Education*. doi:10.1080/03075079.2015.1007944.

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). Los Angeles, CA: Sage.

Taras, M. (2002). Using assessment for learning and learning from assessment. *Assessment & Evaluation in Higher Education*, *27*(6), 501-510.

Wallander, L. (2009). 25 years of factorial surveys in sociology: A review. *Social Science Research*, *38*(3), 505-520.

Whitington, V., Glover, A., & Harley, F. (2004). Preservice early childhood students' perceptions of written feedback on their essays. *Early Child Development and Care*, *174*(4), 321-337.

Williams, K. (2005). Lecturer and first year student (mis)understandings of assessment task verbs: "Mind the gap". *Teaching in Higher Education*, *10*(2), 157-173.

Winstone, N. E., Nash, R. A., Rowntree, J., & Menezes, R. (2015). What do students want most from written feedback information? Distinguishing necessities from luxuries using a budgeting methodology. *Assessment & Evaluation in Higher Education*. doi:10.1080/02602938.2015.1075956.

7 General discussion

7.1 Core findings and contributions

The overall research agenda of this thesis was divided into three parts, each of which the subject of an empirical investigation. Study 1 attempted to establish the significance of fair grading procedures for student attitudes toward their studies as a whole. Specifically, the question was to what extent students' dropout intentions are related to their perceptions of the fairness of grading procedures. Study 2 introduced a shift in perspective by treating justice perceptions as outcomes instead of as predictors of an outcome. The focus of this article was on how justice perceptions are affected by structural characteristics on the level of the university department, namely the department-specific configuration of assessment method and instruction method. Finally, Study 3 added to our knowledge regarding the formation of justice perceptions by exploring the process through which students arrive at a justice judgment by weighing the importance of various properties of a situation against each other. The following paragraphs give a quick summary of the core results of each study as well as a discussion of their contribution to the state of research.

7.1.1 Study 1

The investigation of the relationship between the fairness of grading procedures and dropout intentions provided the first empirical application of the conception of fair grading procedures as developed in this thesis. The central theoretical premise of Study 1 was that students consider matters of procedural justice when contemplating whether or not they should prematurely abandon their studies. The reason for this lies in the connection between the fairness of the grading process and the perceived chances of successfully graduating from university. As long as grading manages to live up to the standards demanded from a fair procedure, students can gauge their chances of success with greater certainty. Conversely, unfair procedures favor the interpretation that students are kept from succeeding by factors over which they have no control.

Regression analyses provided evidence in favor of the theoretical argument that contributes to our understanding of the formation of dropout intentions. The results showed that dropout intentions are affected by perceptions of both validity-related procedural justice and control-related procedural justice. The effect of PJ-V was found to be

stronger compared to the effect of PJ-C. After controlling for satisfaction with academic achievement, PJ-V continued to show statistical significance whereas PJ-C did not. This supports the proposal of treating validity and control-related aspects of procedural justice as separate entities. An additional layer was added to the analyses by showing that the effects of justice perceptions are moderated by the students' social background. Effects of PJ-V were found to be stronger for second generation immigrants. Effects of PJ-C were stronger for both second generation immigrants and students from low-SES families. These effects were present even when controlling for academic achievement, which shows that for students from these backgrounds, having additional security is particularly essential.

This study represents the first attempt to link dropout intentions to the fairness of grading procedures. The approach that was chosen here follows the prevailing strategy of justice research of using justice perceptions as predictors of an outcome. However, it also expands on this strategy by using a stricter focus. Instead of employing a broad operationalization of procedural justice that subsumes procedures as disjunct as grading and attendance policies or even asking about "procedures" in general, the theoretical framework and its empirical implementation were explicitly focused on the procedures used to grade assessments of student performance. This allowed for a clear specification of *why* perceptions of both types of procedural justice were expected to affect dropout intentions. This would not have been possible if procedural justice had been operationalized as a more or less unsystematic amalgamation of different procedures. Admittedly, it is possible to establish a correlation between an unspecific procedure and a specific outcome. This could also work in case of unspecific procedures and dropout intentions. However, the explanatory power of such a result would be limited by the lack of specificity in the conceptualization of its explanatory variables: What are the mechanisms through which the fairness of procedures affects dropout intentions? It is questionable that there is an answer to this question that pertains to "procedures" in general, since different procedures govern the distribution of different resources. Therefore, Study 1 explained dropout intentions through a procedure that affects the distribution of a resource directly related to educational success, i.e., the procedure used to allocate grades to students. It then gave an account of the underlying mechanisms that link these procedures to dropout intentions.

Future research would benefit from using a stricter focus similar to this study. This is not limited to the investigation of student dropout, just as it is not limited to justice

in education. Some of the outcomes investigated by procedural justice research are desirable (e.g. achievement, identification), others are not (e.g. dropout, anti-social behavior). If this research is to have practical relevance, it should also determine ways to promote desirable outcomes and to prevent undesirable ones. The first step in this direction is to identify *which* procedures affect the outcome of interest. Attempting to improve justice perceptions without having a clear sense of which procedures should be targeted is neither efficient nor is it likely to achieve the desired result. The second study of this thesis expanded on this argument. It demonstrates how focusing on well-defined procedures allows us to not only provide a conclusive account of how individuals are affected by their perceptions of the procedures, but also of what determines how these perceptions develop.

7.1.2 Study 2

The finding that unfair grading procedures can have a negative impact on student retention indicates a need to improve the students' justice perceptions. However, this agenda is held back by the fact that presently, there is a lack of empirically grounded knowledge about what causes justice perceptions to go in a certain direction. To take on this issue, Study 2 investigated the role of the academic environment in shaping justice perceptions.

The decision to focus on the institutional level instead of the individual level was inspired by the *justice climate* approach (Li et al. 2015). The central argument is that students in the same department are subject to the same organizational structure and are therefore likely to make similar experiences. Hence, it is critical to consider the context from which justice perceptions emerge. Study 2 attempted to model the organizational structure by assigning a set of concrete properties to each department: the extent to which assessment is either essay-based or exam-based as well as the extent to which instruction is either seminar-based or lecture-based. The basic premise of this research was that each of these methods has certain properties that affect how fair the procedures appear from the students' perspective. Whereas the choice between essays and exams is decisive for the degree to which assessment is standardized, the comparison of seminars and lectures points to substantial differences in how students interact with faculty. Starting from this, Study 2 set out to investigate how these structural differences on the department level reflect on the students' justice perceptions.

The effects of department-level characteristics on individual-level perceptions of PJ-C, PJ-V, and IJ were explored in a series of mixed effects regression models. The results of these models provide valuable insights into how justice perceptions depend on

the favorability of the academic environment. Ratings of PJ-C were found to be higher in departments in which assessment places greater weight on essays rather than exams. It appears that due to its low degree of standardization, the essay format provides more room for student involvement. However, a stronger emphasis on assessment via essays was found to actually decrease ratings of PJ-V. Therefore, the very flexibility that promotes student involvement appears to also favor the impression that the grading process lacks validity. As the openness of the assessment process increases, ambiguity increases as well. This gives additional support for separating control-related from validity-related aspects of procedural justice. The diverging effects of the assessment method on ratings of PJ-C and PJ-V could not have been identified if control and validity had been treated as a single construct. Thus, it is not only hypothetically possible that the grading process can be fair with regard to validity while being unfair with regard to control and vice versa. Rather, the results give empirical evidence that such constellations can be observed in practice. A higher proportion of essays could also be positively linked to perceptions of IJ. Interestingly, the significance of this effect appears to be limited to students from low-SES families. With regard to the instruction method, students in departments in which the ratio between seminars and lectures is in favor of a more student-centered format were shown to give higher ratings of PJ-C and IJ. Similar to the assessment method, feedback-related fairness perceptions of students from low-SES families were shown to depend more strongly on the presence of an environment that promotes student-faculty interactions.

Apart from its contribution to our knowledge about the formation of justice perceptions, Study 2 also provides another clear case for applying a stricter conception of justice and its underlying criteria. Focusing on justice criteria of a specific, well-defined procedure allowed us to identify the structural conditions that influence the procedures as well as the students' perceptions of them. An unspecific conceptualization and operationalization of "procedures" would have undermined the feasibility of this approach. For example, while the implications of the assessment method for the grading process are readily apparent, the same is not necessarily true for other procedures such as attendance policies. Because of that, a lack of specificity with regard to the procedure of interest would have been even more precarious in the context of Study 2 when compared to Study 1.

Altogether, Study 2 presents a good argument for exploring the connection between justice perceptions and the context in which they are observed. While it is readily

apparent that student experiences are marked by enormous differences between departments (Neumann et al. 2002), the implications of these differences for justice perceptions have thus far not been explored (Sabbagh and Resh 2016). This problem is not endemic to research in educational settings, but is characteristic of justice research as a whole. If context is considered at all, like in the justice climate approach, the function assigned to the context appears to be mostly limited to providing an environment in which individuals are assumed to make similar experiences and engage in exchanges regarding these experiences (Li et al. 2015). However, justice climate research has not attempted to describe and analyze concrete properties of the environment that could bring about these similar experiences.

The strategy of focusing on the institutional structure therefore sets the present research apart from previous studies that attended to questions of justice climate. If justice perceptions are seen as reactions to actual procedures, looking at the structural conditions under which the procedures are enacted can provide useful insights, as Study 2 has demonstrated at length. The crucial point here is that through modeling the conditions that affect justice perceptions, we can not only detect discrepancies between departments. Rather, we are also given access to the underlying causes of between-departmental differences, thereby supplying us with clear targets for reducing these differences.

A compelling correlate of this strategy is worth mentioning here. Contrary to previous conceptions of justice climate, the theoretical framework proposed here does not depend on the assumption that individuals in a department exchange their views regarding matters of fairness (Liao and Rupp 2005). While it is still acknowledged that justice perceptions can evolve from exchanges between peers, this thesis established how structural conditions govern individual perceptions. This means that if assessment and instruction are organized in a way that affects the likelihood with which justice-related expectations are met, individuals can arrive at similar perceptions regardless of whether or not they interact with their peers. Exchange between individuals can confirm and reinforce these perceptions, but the basic input comes from the institutional structure in which these relationships are embedded.

7.1.3 Study 3

Study 3 kept the focus on the formation of justice perceptions that was introduced in Study 2. However, the subject was approached from a rather different direction. The first two studies of this thesis operationalized justice perceptions via indirect measures. This

means that the participating students were asked about the extent to which grading procedures and assessment feedback are in accordance with the various criteria used to define the respective justice dimensions. These indirect measures were then used to infer whether or not a student experiences injustice. In Study 3, the task of making the final judgment was transferred to the students themselves. The aim of this study was to show how students utilize various characteristics of a feedback situation to arrive at a judgment of informational justice.

At the center of this study was a factorial survey experiment. The vignettes contained dimensions that describe both the feedback itself as well as the circumstances under which the feedback was given. The research agenda was divided into three steps. First, the goal was to establish how students weigh the importance of different dimensions such as feedback extent and timeliness against each other when forming an opinion about the fairness of the situation. Regression results show that the largest influences on vignette ratings can be attributed to the extensiveness of descriptive feedback and constructive feedback as well as to feedback comprehensibility. In terms of extent, students assigned more weight to constructive feedback than to descriptive feedback. The vignettes also gave information on the timeliness with which feedback is delivered as well as on the mode of delivery. Students were found to have a preference for feedback that comes with little delay and in face-to-face communication instead of in written form. But while the effects of feedback timeliness and mode were highly significant, their impact on ratings of informational justice was rather limited when compared to feedback extent and comprehensibility.

In the next step, the study investigated the moderation of the effect of feedback extent by the grade of the assignment for which feedback is given. The goal was to show how feedback-related expectations are adjusted to the context of the assessment. Here, we saw that the demand for both descriptive and constructive feedback increases as grades decrease. Based on these results, we inferred that a student will feel less injustice after receiving little feedback if they are satisfied with their grade than they would in case of a more troublesome result. The moderation was more pronounced for constructive feedback.

The final step picked up a theme introduced in Study 2 of this thesis: the role of the academic environment. Based on Becher's (1989) classification of academic disciplines, participants were assigned to either hard pure, hard applied, soft pure, or soft applied disciplines. Results showed that in addition to the differences in feedback practices

between disciplines described in Study 2, students also react differently to the feedback they receive. Insufficient explanations of grading criteria were perceived negatively in all fields, but this effect was larger for students from soft pure disciplines in case they received a less-than-good grade.

In addition to enhancing our understanding of students' feedback-related expectations, Study 3 points to some issues that are usually not given too much thought in justice research (admittedly, the first two studies of this thesis were no exception in this regard). While the results of the factorial survey show that the general criteria used to define informational justice are actually relevant for the students' justice perceptions, the criteria are not seen as equally important. This demonstrates that in practice, justice judgments are not formed by simply averaging over a number of criteria. The actual process is certainly more complex than that. Some of this complexity could be captured in the factorial survey experiment: dimensions are weighted against each other, the weight of a dimension is adjusted to the situation, and the judgment is made in light of the students' previous experiences with assessment and feedback in their disciplinary context. These aspects of the decision-making process could not have been captured using conventional, item-based measures.

This raises a rather important question: To what extent do item-based measures of justice perceptions correspond to the actual sense of justice? Based on the fact that the justice criteria chosen in this thesis are deeply grounded in justice theory, we can assume that all of them are somewhat relevant for justice perceptions. But as the results of the factorial survey suggest, the information gathered from item-based measures are best treated as approximations of actual justice perceptions rather than accurate representations. It is to be expected that this is not limited to informational justice, but also pertains to other justice dimensions. Take PJ-C as an example: Can the mere fact that a student reports to have little involvement in the grading process be used as evidence that they feel injustice? What if the student was not involved, but did not *expect* to be involved either? In case of PJ-V, it is safe to assume that students expect that grading should always be free of bias, consistent, and accurate. However, the fact that all criteria should be important does not entail that they are *equally* important. Bias suppression could be valued higher than accuracy, but the opposite could be true just as well. In addition, preferences could depend on the specifics of an assignment, just as they could differ between institutional contexts.

Justice research would be well advised to give more consideration to these matters. One of the central proposals of this thesis was that a stricter focus needs to be applied when investigating justice perceptions – use a well-defined set of justice criteria for the investigation of an equally well-defined procedure. Study 3 shows that arriving at a better understanding of how these criteria relate to each other needs to be added to this agenda. In order for this to succeed, additional factorial survey experiments will be necessary to establish the preferential structure of other justice dimensions as well.

7.2 Implications for future research

A number of possible avenues for future research has already been discussed in the concluding sections of each individual study. But beyond these rather specific points, we can arrive at a more general agenda that can guide subsequent research activities concentrated on justice in education. The strategy deployed in this thesis can serve as a template for such research: First, establish the importance of the fairness of a well-defined procedure. Second, identify factors that determine whether or not this procedure is seen as fair. A crucial advantage of this approach is that it helps us to pinpoint concrete areas in which justice is an issue, which facilitates the search for the causes of these issues.

7.2.1 Beyond grading

First, it is essential to explore justice perceptions of procedures other than grading as well. By designating the fairness of the grading process as the procedure of interest, this thesis was given a very specific focus. The benefits of this strategy have been thoroughly discussed. It ensures that concrete mechanisms can be defined through which justice perceptions affect the students, just as it allows us to determine what affects the justice perceptions themselves. However, the downside of this high degree of specificity is that other potentially important procedures had to be neglected. To give a few examples, procedures like university admission, admission to classes, the scheduling of classes, and the allocation of financial resources to departments could also be targets of justice-related concerns. There is hardly any reason to presume that procedures that are unrelated to each other satisfy the various justice criteria to the same degree. For example, the presence of accurate grading procedures does not necessitate the presence of accurate admission procedures and vice versa. Thus, the mere fact that grading procedures are seen as fair is not enough to conclude that students do not experience injustice in a number of other areas.

If the goal is to provide a learning experience that is generally seen as fair, future research is advised to consider perceptions of the fairness of a wider range of procedures that govern the distribution of important resources. The specification of justice dimensions and their constituent criteria that were used in this thesis to investigate grading procedures can be applied to other procedures as well. For example, the procedures that govern the admission to classes can be judged according to whether students have a voice in the process, the procedures are keeping to strict standards, and provide explanations that ensure transparency. Going from there, the fairness of each dimension can be held to the criteria used in this thesis. The aim should be to introduce a certain level of consistency with regard to the conceptualization and operationalization of justice perceptions. This allows systematic comparisons of perceptions across a wide range of different procedures, which in turn supports the identification of areas where justice-related problems are especially pronounced.

This could be greatly enhanced by also considering the preferential structure that underlies perceptions of the fairness of other procedures. As Study 3 has demonstrated, the importance students assign to different criteria of informational justice is not balanced. From this, it was concluded that additional factorial survey experiments are necessary that explore the preferential structure at the core of perceptions of PJ-C and PJ-V. Again, these investigations should not be limited to the fairness of grading procedures. It is certainly plausible that the position in the preferential hierarchy students assign to a specific justice criterion differs between procedures. From the results of Study 3, we can infer that the timeliness of feedback related to the *assessment process* is of secondary importance to the students. But pending further research, we should not use this as evidence that timeliness is negligible for informational justice in general. Receiving feedback with minimal delay could be seen as highly critical in the context of other procedures.

Arriving at a better understanding of which criteria are decisive for justice perceptions of different procedures would permit a more focused access to improving justice perceptions. If a justice criterion is found to be of great importance for a specific procedure, interventions that target feelings of injustice should give special priority to this criterion. Conversely, attempts to improve an aspect of a procedure that is seen as unimportant are not likely to lead to substantial improvements in justice perceptions, which is why efforts and resources can be directed toward more pressing needs.

7.2.2   Context and structure

Finding that a specific procedure causes justice-related concerns should lead to the question of what causes these concerns. With regard to this, this thesis developed a novel approach to the investigation of what determines whether or not procedures are seen as fair. It did so by focusing on the role of different methods of assessment and instruction in different university departments. Future research should follow this path by devoting more attention to structural conditions that affect individual experiences. The reasoning behind this is pretty straightforward: Without knowing why procedures are seen as unfair, it is hardly possible to combat feelings of injustice.

Expanding on the work done in this thesis, it is to be expected that there are additional factors on the department level that have an impact on perceptions of the fairness of grading procedures. The choice to investigate assessment method and instruction method was motivated by the aim to focus on department-level variables that can be assumed to be relevant across all departments in the sample. Even though the weight placed on each type of assessment and instruction is subject to immense variation, students of all departments will at some point come across some forms of essays, exams, seminars, and lectures. While this ensures that results can be compared across departments, it can also prevent the detection of other influential factors that are only relevant in a subset of departments (or even a single department). For example, justice perceptions could depend on the way laboratory experiments are organized, but this is only relevant in departments where laboratory experiments are part of the curriculum.

Of course, placing the focus on procedures other than the grading process will also require the identification of other structural determinants. The aim should be to arrive at a comprehensive picture of how students' justice related attitudes are affected by the institutional environment. A clear advantage of this approach is its ability to identify concrete structural properties that can be targeted by measures that aim at an improvement of justice perceptions. This provides us with policy implications whose practical relevance is arguably higher than what can be gathered from attempts to explain justice perceptions on grounds of purely individual level factors. To a great extent, changing the academic environment is a matter of available resources. This is without a doubt a highly critical point that could impose major restrictions on possible changes. Nevertheless, it is safe to assume that structural factors such as the organization of assessment and instruc-

tion are still more accessible to interventions than the students' psychological dispositions. While it is highly unlikely that a state can be reached where no student ever experiences injustice, it would be commendable to at least address the structural sources of such feelings in a way that significantly reduces the discrepancies between departments.

Note that the benefits of this approach extend beyond educational settings. Wherever the distribution of resources is governed by rules and regulations on the institutional level, there is a fair chance that differences in rules and regulations result in differences in justice perceptions. Recognizing the importance of the structural properties of institutional environments opens up a vast, unexplored field of compelling research questions. For example, if the perceived fairness of pay raise procedures is of interest, factors like the frequency with which pay raises occur, the average size of pay raises, or the criteria used to decide whose salary is raised by how much could presumably explain differences in justice perceptions between companies (or between departments within companies). The results gathered from such research promise a thorough understanding of the causes behind perceived injustice so that adjustments can be made to structural configurations that are found to be particularly problematic.

7.2.3 Measuring justice perceptions

The last point to be made here concerns the practice of assessing justice perceptions via item-based measures. Establishing the preferential structure of different justice dimensions in different procedures is an important task for which factorial surveys are suited particularly well. Nevertheless, it needs to be emphasized that the knowledge gathered from factorial surveys pertains to hypothetical situations, meaning that we cannot draw any direct inferences regarding *actual* justice perceptions.

However, combining factorial surveys with item-based measures could provide us with an innovative approach for measuring justice perceptions. Take the case of PJ-V as an example. A factorial survey could be set up in which students are presented vignettes that describe an instructor's conduct in the grading process. The vignette dimensions represent bias suppression, consistency, and accuracy. Based on ratings of the fairness of the grading process, it is possible to determine the relative weight students assigns to each dimension. In the next step, these weights can be included in the calculation of a score from item-based measures of the students' own experiences with the same criteria. This ensures that the resulting score accounts for the relative importance of each criterion from the students' perspective, which promises a more accurate estimate of perceived injustice.

To illustrate this: Imagine a factorial survey in which the vignette ratings given by two students indicate that for Student A, consistency is the most important dimension of PJ-V, while it is the least important for Student B. The same study also includes an item-based measure of PJ-V in which both students stated that grading procedures are highly inconsistent. Their ratings of bias suppression and accuracy are identical as well. Without accounting for the results of the factorial survey, both students would be assigned the same score for perceptions of PJ-V. However, in light of the preferences derived from the vignette ratings, this is obviously inaccurate. Since Student A places more weight on consistency, they should feel greater injustice than Student B. By incorporating the weights into the calculations of the scores, we can adjust for these differences. Even though this strategy requires additional efforts, it would help to alleviate some of most salient problems that come with conventional approaches for measuring justice perceptions.

Of course, the applicability of this approach is not confined to educational settings. Indeed, accounting for the fact that, for example, job promotions that are free of bias can be more important to some employees than to others would greatly increase the accuracy with which justice perceptions are measured. This is particularly important for research that aims to analyze consequences of justice perceptions. For example, if two employees give the exact same ratings on a set of items that measure the fairness of a job promotion procedure, we would assume that both experience the same degree of injustice. Therefore, an outcome like job satisfaction should be affected by their justice perceptions to the same extent. However, if the preferential hierarchy of the different justice criteria varies between both individuals so that the procedure is actually seen as highly unfair by one, but rather neutral by the other, the effect on the outcome should not be the same at all. The further observed justice perceptions deviate from actual justice perceptions, the lower the precision of estimates that relate justice perceptions to an outcome, and the greater the risk of drawing inaccurate conclusions from these results. Therefore, it is advised that researchers devote more attention to these issues.

7.3 Conclusion

The introductory chapters to this thesis gave an account of the importance justice has in our daily lives. The idea that things should be fair is so deeply ingrained into our thinking that many times, we cannot help but judge the fairness of events that affect us. Of all things we demand to be fair, grades are among the most salient ones – virtually everyone

who has ever set foot in an educational institution has at one point been on the receiving end of a grading decision. Grades guide us through a significant portion of our education, and wherever grades are given, questions of fairness are likely to follow.

Combining the ideas that grades are important and that important things should be distributed in a fair manner, the logical conclusion should be that grading procedures should be fair. Looking past the simplicity of that statement, it is somewhat surprising that research on questions of procedural justice in education is still in its infancy (Sabbagh and Resh 2016). This is even more pertinent for informational justice, where available research consists of a single study among students in secondary education (Kazemi 2016). The general scarcity of research on justice in education is further compounded by the lack of a consistent strategy for how to conceptualize and operationalize justice perceptions. Building upon the theoretical and empirical foundations of organizational justice research, this thesis set out to mitigate the issues found in previous research on justice in education.

The basic aim was twofold: First, demonstrate the consequences of justice perceptions. Second, demonstrate how these perceptions develop. The first step toward this goal was to define a coherent framework for the investigation of justice perceptions in a higher education context. The main emphasis in building this framework was placed on specificity: Use a clearly delineated set of justice criteria to describe the fairness of an equally well-defined procedure. The focus was set on the fairness of grading procedures (PJ) as well as on the information students are given regarding these procedures (IJ). With respect to procedural justice, this thesis distinguished questions of student involvement in the grading process (PJ-C) from questions regarding the validity of the grading process (PJ-V).

This framework was then applied in a series of empirical studies. The first of these examined students' reactions to the fairness of the grading process. The finding that grading-related justice perceptions are associated with the intention to prematurely abandon one's studies underlines the criticality of the subject. In light of the continuously high numbers of university dropouts in Germany and elsewhere, student perceptions of the assessment process should be given greater consideration. Moving past the repercussions of unfair grading procedures, this thesis therefore made a strong case for giving more thought to the formation of justice perceptions. This is not to be taken as a prompt to abandon research on outcomes of justice perceptions altogether. Instead, research should

follow the strategy devised in this thesis by considering both consequences and antecedents of justice perceptions in an integrated framework. The first foray into this direction addressed the question *why* grading procedures and explanations relating to them are seen as unfair. This was done by exploring the role of structural characteristics of the academic environment. This was followed by an investigation of *how* justice perceptions are formed. Using data collected in a factorial survey experiment, this thesis modeled the complex process through which students combine different properties of a situation and weigh them against each other when forming justice judgments.

In sum, this thesis arrives at a picture of the problems posed by unfair grading procedures while also contributing to the search for practicable remedies. As the discussion in the previous chapter has shown, the implications of these findings extend past the immediate subject of grading procedures. While grades are arguably the most critical resource students receive, they are certainly not the only thing that affects the learning experience. Students are on the receiving end of a great variety of other distributive procedures, and there is a reasonable chance that they expect these procedures to be fair as well. Future research is encouraged to identify other procedures whose fairness is a vital concern to students, and to find ways to address feelings of injustice related to these procedures. A framework for such endeavors was provided in this thesis.

As a concluding remark, it has to be pointed out that attempts to provide a study experience that caters to the justice-related needs of every single student are neither feasible nor likely to succeed. Some students will always take offense at the way grades are given, just as there will always be instructors for whom fairness is not a priority. But even if there is no conceivable way to completely eliminate feelings of injustice, there is definitely a lot of room for improvement.

8 References

Adams, J. S. (1963). Towards an understanding of inequity. *The Journal of Abnormal and Social Psychology*, *67*(5), 422-436.

Adams, J. S. (1965). Inequity in social exchange. *Advances in Experimental Social Psychology*, *2*, 267-299.

Auspurg, K., & Hinz, T. (2015). *Factorial survey experiments*. Thousand Oaks, CA: Sage.

Bayerlein, L. (2014). Students' feedback preferences: How do students react to timely and automatically generated assessment feedback?. *Assessment & Evaluation in Higher Education*, *39*(8), 916-931.

Becher, T. (1989). *Academic tribes and territories: Intellectual enquiry and the cultures of disciplines*. Milton Keynes: SRHE/Open University Press.

Bies, R. J., & Moag, J. S. (1986). Interactional justice: Communication criteria of fairness. *Research on Negotiation in Organizations*, *1*(1), 43-55.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, *5*(1), 7-74.

Bloxham, S., Boyd, P., & Orr, S. (2011). Mark my words: The role of assessment criteria in UK higher education grading practices. *Studies in Higher Education*, *36*(6), 655-670.

Bloxham, S., Hughes, C., & Adie, L. (2015). What's the point of moderation? A discussion of the purposes achieved through contemporary moderation practices. *Assessment & Evaluation in Higher Education*. doi:10.1080/02602938.2015.1039932.

Blüthmann, I., Lepa, S., & Thiel, F. (2008). Studienabbruch und -wechsel in den neuen Bachelorstudiengängen. *Zeitschrift für Erziehungswissenschaft*, *11*(3), 406-429.

Bobocel, D. R., & Gosse, L. (2015). Procedural justice: A historical review and critical analysis. In R. S. Cropanzano & M. L. Ambrose (Eds.), *The Oxford handbook of justice in the workplace* (pp. 51-87). New York, NY: Oxford University Press.

Breen, R., & Goldthorpe, J. H. (1997). Explaining educational differentials towards a formal rational action theory. *Rationality and Society, 9*(3), 275-305.

Carless, D. (2006). Differing perceptions in the feedback process. *Studies in Higher Education*, *31*(2), 219-233.

Chen, C. Y., Chang, H., Hsu, W. C., & Sheen, G. J. (2016). Learning, behaviour and reaction framework: A model for training raters to improve assessment quality. *Assessment & Evaluation in Higher Education*. doi:10.1080/02602938.2016.1180663.

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

Chory-Assad, R. M., & Paulsel, M. L. (2004). Classroom justice: Student aggression and resistance as reactions to perceived unfairness. *Communication Education*, *53*(3), 253-273.

Colquitt, J. A. (2001). On the dimensionality of organizational justice: A construct validation of a measure. *Journal of Applied Psychology*, *86*(3), 386-400.

Colquitt, J. A., & Rodell, J. B. (2011). Justice, trust, and trustworthiness: A longitudinal analysis integrating three theoretical perspectives. *Academy of Management Journal*, *54*(6), 1183-1206.

Colquitt, J. A., & Rodell, J. B. (2015). Measuring justice and fairness. In R. S. Cropanzano & M. L. Ambrose (Eds.), *The Oxford handbook of justice in the workplace* (pp. 187-202). New York, NY: Oxford University Press.

Colquitt, J. A., Scott, B. A., Rodell, J. B., Long, D. M., Zapata, C. P., Conlon, D. E., & Wesson, M. J. (2013). Justice at the millennium, a decade later: A meta-analytic test of social exchange and affect-based perspectives. *Journal of Applied Psychology*, *98*(2), 199-236.

Crisp, B. R. (2007). Is it worth the effort? How feedback influences students' subsequent submission of assessable work. *Assessment & Evaluation in Higher Education*, *32*(5), 571-581.

Cropanzano, R. S., & Folger, R. (1989). Referent cognitions and task decision autonomy: Beyond equity theory. *Journal of Applied Psychology*, *74*(2), 293-299.

Cropanzano, R. S., & Ambrose, M. L. (2015). Organizational justice: Where we have been and where we are going. In R. S. Cropanzano & M. L. Ambrose (Eds.), *The Oxford handbook of justice in the workplace* (pp. 3-13). New York, NY: Oxford University Press.

Cuseo, J. (2007). The empirical case against large class size: Adverse effects on the teaching, learning, and retention of first-year students. *The Journal of Faculty Development*, *21*(1), 5-21.

Daileyl, R. C., & Kirk, D. J. (1992). Distributive and procedural justice as antecedents of job dissatisfaction and intent to turnover. *Human Relations, 45*(3), 305-317.

Deeley, S. J., & Bovill, C. (2015). Staff student partnership in assessment: Enhancing assessment literacy through democratic practices. *Assessment & Evaluation in Higher Education*. doi:10.1080/02602938.2015.1126551.

Degoey, P. (2000). Contagious justice: Exploring the social construction of justice in organizations. *Research in Organizational Behavior*, *22*, 51-102.

Denton, P., Madden, J., Roberts, M., & Rowe, P. (2008). Students' response to traditional and computer-assisted formative feedback: A comparative case study. *British Journal of Educational Technology*, *39*(3), 486-500.

Deutsch, M. (1975). Equity, equality, and need: What determines which value will be used as the basis of distributive justice?. *Journal of Social Issues*, *31*(3), 137-149.

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

Dirks, K. T., & Ferrin, D. L. (2001). The role of trust in organizational settings. *Organization Science*, *12*(4), 450-467.

Esser, H. (1999). *Soziologie: Spezielle Grundlagen Vol. 1, Situationslogik und Handeln*. Frankfurt: Campus.

Evans, C. (2013). Making sense of assessment feedback in higher education. *Review of Educational Research*, *83*(1), 70-120.

Ferguson, P. (2011). Student perceptions of quality feedback in teacher education. *Assessment & Evaluation in Higher Education*, *36*(1), 51-62.

Fox, S., Spector, P. E., & Miles, D. (2001). Counterproductive work behavior (CWB) in response to job stressors and organizational justice: Some mediator and moderator tests for autonomy and emotions. *Journal of Vocational Behavior, 59*(3), 291-309.

Georg, W. (2009). Individual and institutional factors in the tendency to drop out of higher education: A multilevel analysis using data from the Konstanz Student Survey. *Studies in Higher Education, 34*(6), 647-661.

Gibbs, G., & Simpson, C. (2004). Conditions under which assessment supports students' learning. *Journal of Teaching and Learning in Higher Education, 1*(1), 1-30.

Greenberg, J. (1990). Organizational justice: Yesterday, today, and tomorrow. *Journal of Management, 16*(2), 399-432.

Greenberg, J. (1993). The social side of fairness: Interpersonal and informational classes of organizational justice. In R. Cropanzano (Ed.), *Justice in the workplace: Approaching fairness in human resource management* (pp. 79-103). Hillsdale, NJ: Erlbaum.

Hand, L., & Clewes, D. (2000). Marking the difference: An investigation of the criteria used for assessing undergraduate dissertations in a business school. *Assessment & Evaluation in Higher Education, 25*(1), 5-21.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112.

Heublein, U. (2014). Student drop-out from German higher education institutions. *European Journal of Education, 49*(4), 497-513.

Heublein, U., Hutzsch, C., Schreiber, J., Sommer, D., & Besuch, G. (2010). *Ursachen des Studienabbruchs in Bachelor-und in herkömmlichen Studiengängen (HIS: Forum Hochschule 2 – 2010)*. Hannover: HIS.

Higgins, R., Hartley, P., & Skelton, A. (2001). Getting the message across: The problem of communicating assessment feedback. *Teaching in Higher Education, 6*(2), 269-274.

Homans, G. C. (1961). *Social behavior: Its elementary forms*. Oxford: Harcourt.

Hounsell, D., McCune, V., Hounsell, J., & Litjens, J. (2008). The quality of guidance and feedback to students. *Higher Education Research & Development*, *27*(1), 55-67.

Jasso, G. (1978). On the justice of earnings: A new specification of the justice evaluation function. *American Journal of Sociology*, *83*(6), 1398-1419.

Jasso, G. (1980). A new theory of distributive justice. *American Sociological Review*, *45*(1), 3-32.

Jasso, G., & Resh, N. (2002). Exploring the sense of justice about grades. *European Sociological Review, 18*(3), 333-351.

Jones, E. B., & Jackson, J. D. (1990). College grades and labor market rewards. *The Journal of Human Resources*, *25*(2), 253-266.

Kazemi, A. (2016). Examining the interplay of justice perceptions, motivation, and school achievement among secondary school students. *Social Justice Research*, *29*(1), 103-118.

Konovsky, M. A., & Pugh, S. D. (1994). Citizenship behavior and social exchange. *Academy of Management Journal*, *37*(3), 656-669.

Korsgaard, M. A., Schweiger, D. M., & Sapienza, H. J. (1995). Building commitment, attachment, and trust in strategic decision-making teams: The role of procedural justice. *Academy of Management journal*, *38*(1), 60-84.

Lang, V., & Hillmert, S. (2014) (Eds.). *CampusPanel User Handbook V1.1: Documentation for the Student Panel of the ScienceCampus Tuebingen, wave "a"*. Tübingen: Institut für Soziologie. http://www.campuspanel.ifsoz.uni-tuebingen.de/files/campuspanel_userhandbook_v1_1.pdf. Accessed 10 November 2015.

Leventhal, G. S. (1976). The distribution of rewards and resources in groups and organizations. In L. Berkowitz & E. Walster (Eds.), *Advances in experimental social psychology* (Vol. 9, pp. 91-131). New York, NY: Academic Press.

Leventhal, G. S. (1980). What should be done with equity theory? In K. Gergen, M. Greenberg, & R. Willis (Eds.), *Social exchange: Advances in theory and research* (pp. 27-55). New York: Plenum Press.

Li, A., & Cropanzano, R. S. (2009). Fairness at the group level: Justice climate and intraunit justice climate. *Journal of Management*, *35*(3), 564-599.

Li, A., Cropanzano, R. S., & Molina, A. (2015). Fairness at the unit level: Justice climate, justice climate strength, and peer justice. In R. S. Cropanzano & M. L. Ambrose (Eds.), *The Oxford handbook of justice in the workplace* (pp. 137-164). New York, NY: Oxford University Press.

Liao, H., & Rupp, D. E. (2005). The impact of justice climate and justice orientation on work outcomes: A cross-level multifoci framework. *Journal of Applied Psychology*, *90*(2), 242-256.

Liebig, S., & Sauer, C. (2016). Sociology of justice. In C. Sabbagh & M. Schmitt (Eds.), *Handbook of social justice theory and research* (pp. 37-59). New York, NY: Springer.

Liebig, S., Sauer, C., & Friedhoff, S. (2015). Using factorial surveys to study justice perceptions: Five methodological problems of attitudinal justice research. *Social Justice Research*, *28*(4), 415-434.

Lind, E. A., & Van den Bos, K. (2002). When fairness works: Toward a general theory of uncertainty management. In B. M. Staw & R. M. Kramer (Eds.), *Research in organizational behavior* (Vol. 24, pp. 181-223). Boston: Elsevier.

Lind, E. A., & Tyler, T. R. (1988). *The social psychology of procedural justice.* New York: Plenum Press.

Lizzio, A., & Wilson, K. (2008). Feedback on assessment: Students' perceptions of quality and effectiveness. *Assessment & Evaluation in Higher Education*, *33*(3), 263-275.

López-Pastor, V., & Sicilia-Camacho, A. (2015). Formative and shared assessment in higher education. Lessons learned and challenges for the future. *Assessment & Evaluation in Higher Education*. doi:10.1080/02602938.2015.1083535.

Lynch, R., & Hennessy, J. (2015). Learning to earn? The role of performance grades in higher education. *Studies in Higher Education*. doi:10.1080/03075079.2015.1124850.

Maier, G. W., Streicher, B., Jonas, E., & Woschée, R. (2007). Gerechtigkeitseinschätzungen in Organisationen. *Diagnostica 53*(2), 97-108.

McMorran, C., Ragupathi, K., & Luo, S. (2015). Assessment and learning without grades? Motivations and concerns with implementing gradeless learning in higher education. *Assessment & Evaluation in Higher Education*. doi:10.1080/02602938.2015.1114584.

Moorman, R. H. (1991). Relationship between organizational justice and organizational citizenship behaviors: Do fairness perceptions influence employee citizenship?. *Journal of Applied Psychology*, *76*(6), 845-855.

Mossholder, K. W., Bennett, N., & Martin, C. (1998). A multilevel analysis of procedural justice context. *Journal of Organizational Behavior*, *19*(2), 131-141.

Mulliner, E., & Tucker, M. (2015). Feedback on feedback practice: Perceptions of students and academics. *Assessment & Evaluation in Higher Education*. doi:10.1080/02602938.2015.1103365.

Naumann, S. E., & Bennett, N. (2000). A case for procedural justice climate: Development and test of a multilevel model. *Academy of Management Journal*, *43*(5), 881-889.

Nesbit, P. L., & Burton, S. (2006). Student justice perceptions following assignment feedback. *Assessment & Evaluation in Higher Education*, *31*(6), 655-670.

Neumann, R., Parry, S., & Becher, T. (2002). Teaching and learning in their disciplinary contexts: A conceptual analysis. *Studies in Higher Education*, *27*(4), 405-417.

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, *31*(2), 199-218.

Norton, L. S. (1990). Essay-writing: What really counts?. *Higher Education*, *20*(4), 411-442.

Orr, S. (2007). Assessment moderation: Constructing the marks and constructing the students. *Assessment & Evaluation in Higher Education*, *32*(6), 645-656.

Orsmond, P., & Merry, S. (2011). Feedback alignment: Effective and ineffective links between tutors' and students' understanding of coursework feedback. *Assessment & Evaluation in Higher Education*, *36*(2), 125-136.

Pereira, D., Flores, M. A., & Niklasson, L. (2015). Assessment revisited: A review of research in Assessment and Evaluation in Higher Education. *Assessment & Evaluation in Higher Education*. doi:10.1080/02602938.2015.1055233.

Poon, J. M. (2012). Distributive justice, procedural justice, affective commitment, and turnover intention: A mediation–moderation framework. *Journal of Applied Social Psychology*, *42*(6), 1505-1532.

Rawls, J. (1971). *A theory of justice*. Cambridge, MA: Harvard University Press.

Resh, N., & Sabbagh, C. (2014a). Sense of justice in school and civic attitudes. *Social Psychology of Education*, *17*(1), 51-72.

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

Resh, N., & Sabbagh, C. (2016). Justice and education. In C. Sabbagh & M. Schmitt (Eds.), *Handbook of social justice theory and research* (pp. 349-367). New York, NY: Springer.

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review*, *23*(3), 393-404.

Sabbagh, C., & Resh, N. (2016). Unfolding justice research in the realm of education. *Social Justice Research*, *29*(1), 1-13.

Sabbagh, C. & Schmitt, M. (2016). Past, present, and future of social justice theory and research. In C. Sabbagh & M. Schmitt (Eds.), *Handbook of social justice theory and research* (pp. 1-11). New York, NY: Springer.

Sabbagh, C., Resh, N., Mor, M., & Vanhuysse, P. (2006). Spheres of justice within schools: Reflections and evidence on the distribution of educational goods. *Social Psychology of Education*, *9*(2), 97-118.

Schoorman, F. D., Mayer, R. C., & Davis, J. H. (2007). An integrative model of organizational trust: Past, present, and future. *Academy of Management Review*, *32*(2), 344-354.

Severiens, S. E., & Schmidt, H. G. (2009). Academic and social integration and study progress in problem based learning. *Higher Education*, *58*(1), 59-69.

Severiens, S. E., Meeuwisse, M., & Born, M. (2015). Student experience and academic success: Comparing a student-centred and a lecture-based course programme. *Higher Education*, *70*(1) 1-17.

Shapiro, D. L., Buttner, E. H., & Barry, B. (1994). Explanations: What factors enhance their perceived adequacy?. *Organizational Behavior and Human Decision Processes*, *58*(3), 346-368.

Simons, T., & Roberson, Q. (2003). Why managers should care about fairness: The effects of aggregate justice perceptions on organizational outcomes. *Journal of Applied Psychology*, *88*(3), 432-433.

Skarlicki, D. P., & Folger, R. (1997). Retaliation in the workplace: The roles of distributive, procedural, and interactional justice. *Journal of Applied Psychology*, *82*(3), 434-443.

Small, F., & Attree, K. (2015). Undergraduate student responses to feedback: Expectations and experiences. *Studies in Higher Education*. doi:10.1080/03075079.2015.1007944.

Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics*, *87*(3), 355-374.

Tannock, S. (2015). No grades in higher education now! Revisiting the place of graded assessment in the reimagination of the public university. *Studies in Higher Education*. doi:10.1080/03075079.2015.1092131.

Thibaut, J. W., & Walker, L. (1975). *Procedural justice: A psychological analysis*. Hillsdale, NJ: Erlbaum.

Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research*, *45*(1), 89-125.

Tyler, T. R. (2000). Social justice: Outcome and procedure. *International Journal of Psychology*, *35*(2), 117-125.

Tyler, T. R., & Lind, E. A. (1992). A relational model of authority in groups. *Advances in Experimental Social Psychology*, *25*, 115-191.

Vermunt, R., & Steensma, H. (2016). Procedural justice. In C. Sabbagh & M. Schmitt (Eds.), *Handbook of social justice theory and research* (pp. 219-236). New York, NY: Springer.

Winstone, N. E., Nash, R. A., Rowntree, J., & Parker, M. (2016). "It'd be useful, but I wouldn't use it": Barriers to university students' feedback seeking and recipience. *Studies in Higher Education*. doi:10.1080/03075079.2015.1130032.

Wolfe, E. W. (2004). Identifying rater effects using latent trait models. *Psychology Science*, *46*(1), 35-51.