

Student perceptions of the fairness of grading procedures: a multilevel investigation of the role of the academic environment

Roland Burger¹

Published online: 11 August 2016
© Springer Science+Business Media Dordrecht 2016

Abstract The purpose of this study is to examine the effects of assessment method (essays vs. examinations) 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 1549 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 examinations 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.

Keywords Procedural justice · Informational justice · Academic environment · Assessment method · Instruction method

✉ Roland Burger
roland.burger@uni-tuebingen.de

¹ Department of Sociology, University of Tübingen, Wilhelmstr. 36, 72074 Tübingen, Germany

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 (Dawson 2015; Chen et al. 2016; Bloxham et al. 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 (Naumann and Bennett 2000; Mossholder et al. 1998; 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 (Ramsden 1979; Entwistle and Tait 1990; Neumann 2001). As a result, student experiences vary greatly between departments, which is evident in a number of important outcomes. A large body of the literature has shown that the academic environment influences perceptions of the assessment process (Pereira et al. 2016; Flores et al. 2015; Parpala et al. 2010; Sun and Richardson 2016), academic achievement (Simpson 2015; Godor 2016; Brint et al. 2012), as well as student–faculty interactions (Kim and Sax 2014; Cuseo 2007; 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. Multilevel 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 interdepartmental 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

examinations 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 (Tinto 1993; Pascarella et al. 2004; Ostrove and Long 2007; Rubin 2012). Therefore, special attention will be devoted to the situation of low-SES students.

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* (Thibaut and Walker 1975; Leventhal 1980) 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; they have to be communicated in a timely manner (Colquitt 2001).

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 (Neumann et al. 2002; Lindblom-Ylänne et al. 2006). 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.

Assessment method

While assessment can take place in a variety of other formats such as peer assessment (Topping 1998; Ashenafi 2015), self-assessment (Orsmond and Merry 2015), and portfolio assessment (Dysthe and Engelsen 2011), this study contrasts *essays* and *examinations* as these formats remain by far the most prevalent in the institution studied here. Essays and examinations represent two rather different approaches to measuring student performance, which has consequences for student attitudes toward the assessment process (MacLellan 2001; Scouler 1998; Flores et al. 2015). 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 examinations, 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 examinations 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 examinations 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 examinations is usually shorter than for essays. In many cases, feedback for examinations 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 examinations. 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 lead to the hypothesis that ratings of IJ are higher in departments where assessment is more essay-based (Hypothesis 1_{IJ}).

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_{PJ-C} and Hypothesis 2_{IJ}).

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 (Pascarella et al. 2004; Ostrove and Long 2007). This includes greater insecurities in interactions with faculty and peers (Bourdieu 1986; Tinto 1993; Kim and Sax 2009). 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 examination-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 examination-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 examinations 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 that 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 1.

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

Table 1 Research hypotheses

H1 _{PJ-C}	A higher proportion of essays relative to examinations increases ratings of PJ-C
H1 _{PJ-V}	A higher proportion of essays relative to examinations decreases ratings of PJ-V
H1 _{IJ}	A higher proportion of essays relative to examinations 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

Table 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 ^a	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			

^a In German universities, the level of study is expressed in terms of semesters instead of year of study

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 2.

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

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 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, preenrollment information regarding the study program, and digital media use by faculty. Variables for satisfaction with achievement and preenrollment information are factor scores derived from a CFA using three individual items, respectively (see Table 3).

Department-level predictors

Assessment method: The item measuring the proportion of essays relative to examinations in a department is based on CampusPanel data. Participants were asked how many essays and examinations 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 examinations students have written, the category "exam" subsumes different formats such as multiple-choice questionnaires and open-ended questions. Information on the *last* written examination was available for a small subset of the data. Among these, 57.3 % stated that their last examination 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

Table 3 Scales, reliabilities and descriptive statistics

Scale	Mean	SD	Min	Max
PJ-C (Cronbach's $\alpha = .677$) (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 ($\alpha = .753$) (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 ($\alpha = .897$) (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 ($\alpha = .798$) (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
Preenrollment information regarding program ($\alpha = .775$) (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

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 nonteaching professions.

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).

Results

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 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 ($p < .000$). 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 4 Multilevel regression models, L1 predictors

Level and variable	PJ-C		PJ-V		IJ	
	M1	M2	M1	M2	M1	M2
Level 1						
Intercept	.125** (.044)	.246*** (.054)	-.019 (.040)	.054 (.053)	.070 (.043)	.199*** (.055)
SES (ISEI)		.004 (.014)		.024+ (.014)		.013 (.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, preenrollment information, digital media use by faculty

*** $p < .001$; ** $p < .01$; * $p < .05$; + $p < .1$

Table 5 Multilevel regression models, L1 and L2 predictors

Level and variable	PJ-C	PJ-V	IJ
	M3	M3	M3
Level 1			
Intercept	.292*** (.054)	.167** (.056)	.278*** (.062)
SES (ISEI)	.005 (.014)	.024 (.014)	.015 (.014)
Standard errors in parentheses			
Additional L1 controls: immigrant background, gender, no. of semesters studied, satisfaction with academic achievement, preenrollment information, digital media use by faculty			
Level 2			
Proportion essays	.085*** (.022)	-.052* (.024)	.028 (.026)
Proportion seminars	.036* (.017)	.025 (.018)	.044* (.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

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

* $p < .05$; + $p < .1$

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). 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.

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 examinations 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 examination-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.

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 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 $H3_{IJ}$, which stated that the effect of the assessment method on IJ is larger for low-SES students.

For ease of understanding the interaction between assessment method and SES on IJ, conditional marginal effects are plotted in Fig. 1. Here, we see effects of a 10 % increase

Table 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*** (.053)	.263*** (.061)	.258*** (.061)	.253*** (.061)
SES (ISEI)	.011 (.016)	.010 (.016)	.013 (.016)	.012 (.016)
Level 2				
Proportion essays	.084*** (.022)	.024 (.025)	.033 (.026)	.020 (.025)
Proportion seminars	.039* (.017)	.033 ⁺ (.019)	.031 (.020)	.049* (.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, preenrollment
 information, digital media use by
 faculty
 Additional L2 controls: student-
 faculty ratio, type of degree
 program
 *** $p < .001$; ** $p < .01$;
 * $p < .05$; ⁺ $p < .1$

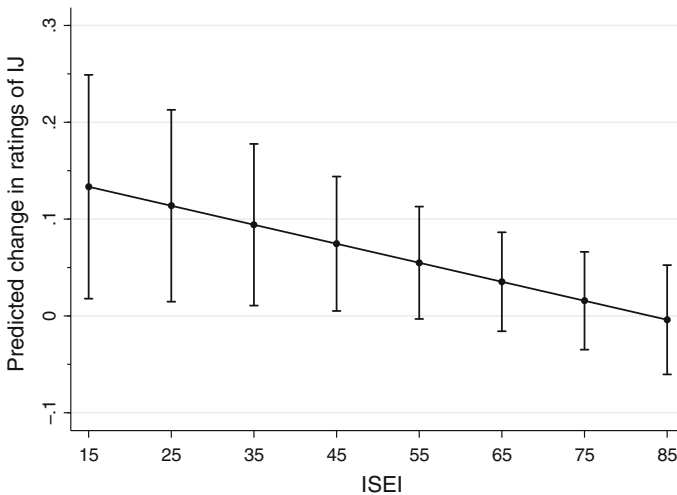


Fig. 1 Cross-level interaction assessment method × SES on IJ. Predicted change in ratings of IJ for a 10 % increase in the proportion of essays for different levels of SES. 95 % confidence bands

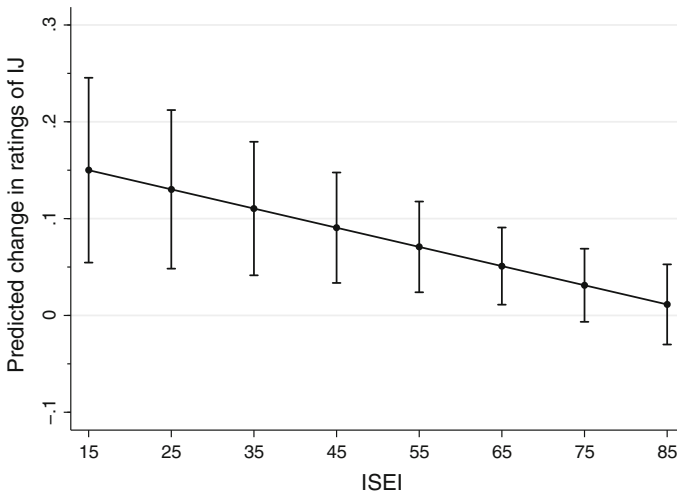


Fig. 2 Cross-level interaction instruction method \times SES on IJ. Predicted change in ratings of IJ for a 10 % increase in the proportion of seminars for different levels of SES. 95 % confidence bands

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.

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_{IJ}. This interaction is visualized in Fig. 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.

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 examinations. 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 where the assessment process is more essay-based as opposed to examination-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. Interdepartmental 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, interdepartmental 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 examination. 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 examinations 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, examinations, seminars, and lectures are rather broad categories that subsume a range of approaches to assessment and teaching. Assessment via examination can take place in many different formats, from multiple-choice questionnaires to open-ended questions. One could argue that negative effects of the examination format on perceptions of PJ-C should be lower in case of open questions than for multiple-choice. Likewise, advantages of examinations 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 (Biglan 1973; Becher 1989; 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.

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:10.1080/02602938.2015.11100711.
- 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 November 10, 2015.
- Burger, R., & Gross, M. (2016). Gerechtigkeit und Studienabbruch. Die Rolle der wahrgenommenen Fairness von Benotungsverfahren bei der Entstehung von Abbruchsintentionen. *Zeitschrift für Erziehungswissenschaft*,. 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 November 10, 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. (Eds.). (2014). *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 November 10, 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., 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.
- Severiens, S. E., & Schmidt, H. G. (2009). Academic and social integration and study progress in problem based learning. *Higher Education*, 58(1), 59–69.
- 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.
- 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.