



<http://www.diva-portal.org>

Postprint

This is the accepted version of a paper published in . This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination.

Citation for the original published paper (version of record):

Beier, S., Eib, C., Oehmann, V., Fiedler, P., Fiedler, K. (2014)  
Influence of judges' behaviors on perceived procedural justice  
*Journal of applied social psychology*, 44(1): 46-59  
<https://doi.org/10.1111/jasp.12199>

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-351145>

Running Head: Judges' behavior and perceived procedural justice

Influence of judges' behavior on perceived procedural justice

Susanne Beier

University of Heidelberg

Constanze Eib

Stockholm University

Verena Oehmann

University of Heidelberg

Peter Fiedler

University of Heidelberg

Klaus Fiedler

University of Heidelberg

*Author Note:* Correspondence regarding this paper should be addressed to Susanne Beier, Psychologisches Institut, Universitaet Heidelberg, Hauptstrasse 47-51, 69117 Heidelberg, Germany. Fax: + 49 / 6221 / 547318. Email:

Susanne.beier@psychologie.uni-heidelberg.de.

*Acknowledgments:* For helpful comments we thank Mandy Hütter, John Pacia, and Jens Hasche. We also thank all the research assistants for their help with the data collection.

### Abstract

The influence of judges' behavior on procedural justice was analyzed in a field study, observing the judges' behavior during  $n = 129$  trials and assessing the defendants' and the audiences' justice perceptions. The observed judicial behavior was unrelated to the defendants' justice perceptions. However, the more respectful the judge treated the defendants, the fairer the audience perceived the trial. In general, the effect size of the relationship between observational measures and subjective justice ratings was small in comparison to the relationship within defendants' or audiences' ratings. There were striking differences in the justice perception between the two data sources, namely defendants and audience. Thus, the source matters and, to avoid a same-source bias, should be taken into account when analyzing justice perceptions.

*Keywords:* procedural justice, judges' behavior, defendants, criminal court, same source bias.

## Introduction

The jurisdiction by criminal courts is one major part of the judiciary branch in a democratic society. The courts' judgments that are announced in the name of the people should be supported by the majority of citizens to be legitimate (Bierbrauer & Klinger, 2008). One mean to gain this support of the society is a fair decision making process in court and a fair verdict. Especially in criminal trials that might lead to harsh sentences that constitute a severe invasion in individual rights, justice is a highly valued good. To assure the commitment of the society and the convicts, even for these severe sentences, the decision making process and the verdict should be fair. In cases where there is no consensus about the verdicts, the only way to assure the acceptance of the judicial institutions by the people is a fair decision making procedure.

The present study focused on a key trial aspect that might be crucial for the defendants' procedural justice perception: the behavior of the judge conducting the trial. More precisely, the present study analyzed the relationship between observed judicial behavior and procedural justice ratings by defendants and the audience.

## Perceived justice

Common to all humans is an inherent need for fairness and justice (Bierbrauer & Klinger, 2008). Early research to understand this core human need focused on the justice of decision outcomes or the distribution of resources termed distributive justice (Tyler, 2000). The distribution of resources is perceived as fair when the allocation is in line with internal norms like the equality or the equity norm with outcomes that are equal or proportional to contributions respectively (Adams, 1965; Deutsch, 1975). Thus, the core assumption of early justice research was that outcomes are the main determinant of perceived justice. However, this perspective of primary result-oriented humans could not be sustained because people's justice perceptions are largely shaped by the way in which decisions are made (for a review see Tyler, 2000). The first systematic empirical work focusing on justice due to the procedures used to make a decision, termed procedural justice (Tyler, 1984), was conducted by Thibaut and Walker (1975). These authors examined the influence of different key procedural characteristic on perceived justice and came to the conclusion that differences in fairness perception are due to differences in perceived control. They distinguished between process and decision control with the former referring to the possibility and the extent to voice one's opinion and the latter to the possibility and the extent of

influence on the final outcome. In their postulated control-oriented model for perceived justice these authors assumed that litigants are not interested in process control per se but just as a mean to gain more favorable results (Tyler & Lind, 1992). Stipulated research, however, suggested a modification of this basic assumption conceptualizing procedural justice as an instrument to attain more favorable outcomes (Lind, 1995). Even negative outcomes were perceived as fairer when the litigant had a subjective experience of control during the trial (Lind & Tyler, 1988). Next to control, additional characteristics of fair procedures were postulated in subsequent research. Leventhal (1980) specified that fair procedures are characterized by adherence to six rules: 1) consistency, 2) bias suppression, 3) accuracy of information, 4) correctability, 5) representativeness and 6) ethicality. These rules were proposed for the fairness perception in non-legal settings. However, they were not based on a strong theoretical foundation (Blader & Tyler, 2003) and were assumed to be valued due to being helpful to achieve an external goal and thus not for the procedure itself (Tyler & Lind, 1992).

Going beyond this instrumental approach the Group Value Theory (Lind & Tyler, 1988) assumes that procedures are valued because they convey information about the relationship with groups one belongs to and authorities representing these groups. Decision making procedures are assumed to represent symbols of group values, with society for example constituting the respective group in court cases. According to this theory, essential for the litigants are the court cases' implications concerning their feeling of embeddedness within society as well as being treated with respect by the societies' institutions. Thus, a trial is perceived as fair if a person gets the impression of being a full-fledged member of the society (Tyler & Lind, 1992). Based on the Group Value Theory and following up on some of Leventhal's (1980) fairness rules (bias suppression), three factors were postulated that shape the procedural justice perception: 1) neutrality of the decision maker, 2) trustworthiness of the decision maker and 3) standing within a group that is inferred by procedures used by the decision maker. Thus, all three factors are associated with the relationship between the defendant and the authority representing the society (Tyler, 1989).

Similar to the proposed factor standing within a group, which is mostly communicated through a polite and dignified treatment (Tyler & Lind, 1992), in organizational justice research a third type of justice has been proposed with interactional justice (Bies & Moag, 1986). This type was again divided in

informational justice, referring to the quality and quantity of information given to explain the application of a specific procedure, and interpersonal justice, referring to the extent of being treated with respect and dignity (Greenberg, 1993). Together with procedural and distributive justice these two types were incorporated in a four factor conceptualization of organizational justice (Colquitt, 2001; Streicher, Jonas, Maier, Frey, Woschée, & Waßmer, 2008) that is supported by meta-analytical evidence (Colquitt, Conlon, Wesson, Porter, & Ng, 2001). However, in court where the trial and thus the application of a procedure are largely shaped by the treatment and information provided by the judge, informational and interpersonal justice might rather be considered as subtypes of procedural justice. In line with this notion, Blader and Tyler (2003) summarized in their procedural justice model the aspects concerning the concrete treatment by an authority as informal quality of treatment that influences justice perception. They contrasted them with formal influences that refer to official rules or laws regulating the treatment by an authority. In addition, they proposed that justice perception is not only influenced by formal and informal aspects of the treatment but also by formal and informal aspects of the decision making. In court these informal aspects of the decision making refer to how a particular judge reached his/ her decision. For the defendants, however, the informal aspects of decision making mostly are accessible through the information provided by the judge and indirectly by the judge's treatment during the trial. Thus, both these informal components (informal quality of the treatment and of the decision making) influencing justice perception in court might mainly refer to the interpersonal treatment by the judge making the judges' treatment during a trial essential for justice perception.

Empirical research yielded support for the following aspects of treatment as main factors contributing to fairness perception. First, when people believe that they are actively involved in the decision making process, with an opportunity to state their own point of view, procedures are perceived as more just. This finding was replicated often since first attained by Thibaut and Walker (1975). Even without changes in the outcome, procedures are experienced as fairer with the opportunity to state one's point of view (Tyler, 2000). Second, there was empirical evidence for the postulated neutrality factor in that perceived honesty, objectivity, and impartiality of the decision maker were positively related to the procedural justice perception. In line with that, procedures subjectively experienced as fair were perceived as being based on rules

and facts (Tyler, 2000). Third, summarizable as Tyler's (1989) postulated trustworthiness factor or factors related to the interpersonal quality of the treatment, the experienced compassion and interest of the decision maker and the inferred motive behind the decision maker's actions are positively related to procedural justice as well (Tyler, 2000). Fourth, the last empirically identified factor influencing the procedural justice perception is the experience of being treated with respect and dignity and thus also refers to an aspect of the quality of the interpersonal treatment. The more respectful a treatment is perceived, the more just the procedures are experienced (Tyler, 2000).

### Judges' Behavior

Judges in German courts are the major judicial authority the defendants are confronted with. Therefore, they play a crucial role for the justice impressions of a defendant. The chief judge conducts the trial, examines the defendant, collects the evidence, which includes the examination of witnesses, and is responsible for the maintenance of order during the trial. Subsequent to the judge, the prosecutor and the attorney can address questions to the defendant and the witnesses. However, the judges' examination of the defendant and the witnesses covers all aspects relevant for the crime the defendant is charged with. Thus, often there is no need for further questions by the prosecutor and the attorney. Especially the prosecutors often waive their right to ask questions and therefore the judge often appears as the main speaker in a trial shaping the examinations and the entire procedure (Wolff & Müller, 1997). Additional to these tasks during the conduction of a trial, it is the judge who decides about a sentence at the end of a trial.

Research on the judges' behavior during criminal trials is very rare. Especially research conducted in Germany, that is most relevant for the present study due to differences in the legal system between different countries, is scarce. Two studies on the impression formed by juvenile prisoners about the judge in their trial indicated that the judges' behavior was perceived as rude in about one third of the cases (Eilsberger 1969; Haller & Machura, 1995). However, these findings are based on surveys with inmates reporting about the trials that led to their imprisonment. These trials were conducted at a different time for each questioned person with the possibility of long time periods between the trial and the questioning. Therefore, the precision of the inmates' ratings of the judges' behavior is questionable due to limited

memory for the judges' actual behavior. Also, their stated impression of the judge might have been influenced by the experiences in jail, by previous encounters with the legal system (Haller, Machura, & Bierhoff, 1995), or by the fact that they blame the judge for being in prison and therefore devalue him/ her. None of these influences on the inmates' ratings was controlled for in both of the studies.

A study that included an online registration of the judges' behavior during trials was conducted by Tausch and Langer (1971). These authors assessed the linguistic behavior of 18 District and County judges toward the defendants at two cases per judge. The judges' linguistic behavior contained disrespectful and discouraging expressions almost with the same frequency that it contained respectful and encouraging expressions. However, this study and the reported data on the judges' behavior might no longer be representative for current judges since the study was conducted three decades ago. Because of the reported age effect with younger judges treating the defendants more respectfully than older ones and assuming that this effect might actually be a cohort instead of an age effect, the generalizability of the findings on current judges' behavior is hard to justify.

#### The present study

Previous research on procedural justice emphasized the importance of the litigants' perception of the treatment in court for their procedural justice perception (Tyler, 2000). Aspects of trials that were perceived as fair were measured via subjective rating. The standard procedure used in previous research was to let litigants rate the fairness as well as different aspects of a procedure. Then the relationship between these two classes of variables was assessed to infer aspects that characterize a fair procedure. However, the measures for both classes of variables stemmed from the same source. Thus, it is possible that the inferred relationship between aspects of trials and perceived fairness was inflated due to the same-source bias (Colquitt et al., 2001). Hence, one goal of the present research was to measure the justice perceptions and trial characteristics independently in order to assess the relationship between these classes of variables. The trial characteristics were measured via independent observation and their relationship with subjective justice ratings was assessed. Additionally, to estimate the influence of a potential same-source bias in previous research, subjective ratings of trial characteristics by the defendants as well as by people from the audience, as another independent source of data, were also collected

in the present study. Thus, with a multi-source approach we attempt to overcome previously criticized shortcomings of the procedural justice research. More precisely, we wanted to address the shortcoming of a same-source bias (MacCoun, 2005) and the lack of observational data (MacCoun, 2005).

Therefore, we analyzed two competing hypotheses in the present study: If the procedural justice judgments by the defendants are based on their actual, objective treatment in court, then there should be a relation between the judges' behavior measured via observation and the defendants' procedural justice ratings. However, if the previously established relationship between trial characteristics and the defendants' justice impressions mainly resulted from using the same data source for these measures, the relation between the judges' behavior and the defendants' justice ratings should be small.

In sum, the present research focused on the influence of the informal quality of the treatment in court by the judge (Blader & Tyler, 2003) on the justice perception with both these classes of variables assessed independently from each other. Via observation, we measured the informal treatment and analyzed its relation to justice perceptions gained from two different data sources: defendants and audience.

## Method

### *Subjects and design*

The study was conducted at multiple courts in the county of the District court of Heidelberg.

*Sample of judges.* Thirteen criminal judges participated voluntarily in the study (six male judges). The judges were between 34 and 57 years old at the beginning of the data collection for the study ( $M = 49.92$ ,  $SD = 8.31$ ).

*Sample of defendants.* The data were collected from December 2008 to July 2009. In total 129 criminal trials were visited from the beginning to the pronouncement of the sentence. In seven cases, there was more than one defendant per trial (six cases with two defendants, one case with three defendants).<sup>1</sup> Of 137 defendants approached after the pronouncement of the judgment, 87 were willing to participate in the study and completed a questionnaire subsequent to the trial (participation rate of 63.5 percent). If defendants refused to participate they mainly did so either because they had a limited amount of time, or because they had little interest in the study.

The defendants willing to participate in the study, were between 16 and 72 years old ( $M = 33.95$ ,  $SD = 14.05$ ).<sup>2</sup> Ten defendants were female. The defendants had between zero and 16 prior criminal records ( $M = 2.87$ ,  $SD = 3.76$ ). The trials in which data of a defendant were collected took between 5 to 690 minutes ( $M = 114.08$ ,  $SD = 112.62$ ).<sup>3</sup> In 31 of these 87 cases a jury was present (35.6 percent), in 15 cases one additional judge was present next to the chief judge (12.1 percent), 23 of the 87 cases were conducted at a District court (26.4 percent), 20 at a juvenile court (23.0 percent) and 16 cases were appellate proceedings (18.4 percent). Data was collected at 2 to 10 trials per judge ( $M = 6.69$ ,  $SD = 2.63$ ). The variety of crimes ranged from minor to severe criminality (e.g., fare evasion, armed robbery). In comparison to the official statistic on condemned persons in Germany (Statistisches Bundesamt, 2009), the present studies' sample of crimes was characterized by a slight overrepresentation of severe crime. Most common crimes in the nationwide statistic of the year 2009 on condemned persons were traffic offense, fraud, burglary/ defalcation and crimes against the physical integrity. In this study's sample, these crimes were frequent too. However, traffic offenses were represented comparatively infrequently in the present study and robbery as well as armed robbery and crimes against sexual self-determination were comparatively more frequent (see Table 1). This overrepresentation of severe crimes might be due to the fact that many cases dealing with minor crimes are settled without a trial in Germany, for example by using penalty order procedures.

*Sample of people from the audience.* In 79 trials people sitting in the audience were willing to participate in the study. In 42.65 percent of these trials more than one person from the audience participated in the present study. On average  $M = 2.07$  ( $SD = 1.74$ ) persons filled out the questionnaire. 27.91 percent of interviewed audience members indicated that they knew the defendant.<sup>4</sup>

*Sample of Observers.* As independent observers served eight psychology students and two students with different majors. Two observers were male. Each trial was observed by two observers.

### *Measures*

*Measures of the judges' behavior.* The judges' behavior toward the defendants during the examination was rated by two independent observers using a coding scheme that was constructed based on research findings concerning aspects of trials perceived as

fair and on communication in court (for a list of all coded variables see Table 2). Given the well documented voice effect in procedural justice research (Tyler, 2000), a rating of the defendants' opportunity to state their view was included in the coding scheme. To represent the well-established positive relationship between the impression of being treated respectfully and perceived procedural justice in the coding scheme, the following aspects of the judge's behavior were coded during observation: the judge's tone, a frequency impression of interruption by the judge and specifically interruptions while the defendant was talking about emotions and a frequency impression of disrespectful comments by the judge. To capture the judge's trustworthiness with the coding scheme, a rating of the judge's compassion, empathy, perspective-taking and interest was conducted as well as a frequency impression rating of questions on emotions and perception of emotional meaning. Compassion was defined as resonating with the emotions of another person (Binting, 2004). Based on the cognitive approach of empathy, this construct was defined as putting oneself in the position of another person (Mischo, 2003). In addition, research findings in the field of communication psychology suggest that active listening signals the conversation partner that he/ she is listened to thoroughly (Montada & Kals, 2007). Behavior indicating active listening might be related to perceived procedural justice. Thus a frequency impression of eye-contact, expressions of comprehension (e.g., nodding, affirmation) and paraphrasing by the judge were included in the coding scheme to represent active listening. To capture the specific communication situation in court, frequency impression ratings of the use of special legal terminology, the use of complex language that was hard to comprehend and direct reference to the wording of the law or specific clauses by the judge were included in the coding scheme (Aronsson, Jönsson, & Linell, 1987). Because of an asymmetrical communication structure between the judge and the defendant in court with an information deficit of the defendant, a rating of the comprehensibility of the questions by the judge was also included in the coding scheme (Aronsson et al., 1987). In addition to ratings that were included because of research findings in the field of procedural justice or communicational psychology, ratings previously used by Tausch and Langer (1971) were added to the coding scheme. These were an impression rating of how often the judge used the word please, how often the judge used commands and how often the judge's expressions were reversible in the sense that the defendant could have used the same expression, previously used by the judge, without being offensive. All

described ratings were assessed on a five point scale with the labels *very seldom* – *seldom* – *occasionally* – *often* – *very often* (Rohrman, 1978).

It was assumed that talking about emotions and addressing them is not necessary and helpful in order to establish the truth in every trial. To not artificially assign low values for variables related to emotions in cases where addressing them would be rather inappropriate or, in the worst case, even hinder establishing the truth, additionally to the frequency impression ratings regarding the actual behavior of the judge a necessity rating of addressing the issue of emotions to get a full understanding of the crime and its consequences was included in the observational scheme. For example, it was rated as rather necessary to address the feelings in a case of a burglary of a homeless person striving to survive but rather not necessary of the defendant in a case of burglary that was committed only for materialistic reasons. If this necessity rating was smaller than the midpoint of 2.5 on a 5 point scale averaged over the two observers, then all other ratings in the coding scheme related to emotions were replaced with missing values (ratings of interruption while the defendant was talking about emotions, compassion, empathy, questions on emotions, perception of emotional meaning). This procedure has the disadvantage of a loss of power due to the replacements with missing values. However, considering emotions only in cases when, judging by the content of the committed crime and trial, it was appropriate and necessary to address emotions was given a priority. This procedure allows differentiating between cases where the judge did not address emotions even though it would have helped to establish the truth versus cases where the judge did not address emotions because, by the nature of the case, it would not have helped to establish the truth in the first place.

For rating the judge's behavior the observers kept track of the behavior patterns' frequency by making tally sheets. These sheets served as orientations for the final frequency impression rating of a behavior's occurrence in relation to the duration at the end of the examination. Using tally sheets should enhance the ratings' accuracy because the judge's behavior and the ratings are more closely connected to each other (Amelang, Schmidt-Atzert, Fydrich, & Moosbrugger, 2006).

In multiple training sessions with video material of re-enacted court trials a consistent use of the coding scheme over the observers was practiced. During the trials not only the behavior during the defendants' examinations was rated, but also the behavior during the victims' and witnesses' examination to gain a larger database

for the assessment of the inter-rater reliability of the coding scheme. For the inter-rater reliability the effective inter-rater reliability was computed (Rosenthal, 1987). Due to a low effective inter-rater reliability ( $R < .60$ ) one variable was excluded from further analyses (comprehensibility of questions by the judge; for inter-rater reliabilities of all variables see Table 2).

In addition to coding these specific judges' behaviors during the examinations, a global rating of the judge at the end of the trial was conducted on items previously used in the study by Tausch and Langer (1971). These bipolar five point scale items had the endpoints *unappreciative/ appreciative*, *unfriendly/ friendly*, *cold-hearted/ warm-hearted*, *impersonal/ personal*, *criticizing/ accepting*, *relentless/ compliant*, *impatient/ patient*. The global rating also included four items that were answered by the defendants too (rating of the judge's courtesy, overall performance of duties, honesty, and sympathy). The effective inter-rater reliability was larger than  $R = .60$  for all items except for the rating of the judge's honesty (all inter-rater reliabilities of the global rating of the judge see Table 3). This item was excluded from further analysis.

*Defendants' Questionnaire.* After the trial the defendants were interviewed with a questionnaire previously used by Haller and Machura (1995). For their questionnaire these authors translated items from Tyler (1984) including items on the absolute outcome of the trial, the outcome relative to expectations and relative to the outcome of others, the satisfaction with the outcome, the distributive as well as the procedural justice, six items on the evaluation of the judges' behavior, items on the role of the jury (if present) during the trial, an item on the fairness of German courts in general as well as ratings of key procedural aspects (adequacy of the courts information as basis for a judgment, time the court took to consider the case, opportunity to present evidence, impartiality of the court, degree to which the court took the evidence into account, equality of weighting the evidence, fairness of the procedural rules).

All questions were completed using a nine-point scale with "1" representing the highest value of the questioned constructs and "9" the lowest. The items on the judges' evaluation were averaged ( $\alpha = .94$ ) resulting in a summary measure for the defendants' impression of the judge. For all statistical analyses the variables were recoded, so that higher numerical values indicate higher values on the measured constructs.

*Audience's Questionnaire.* The questionnaire for the lay people sitting in the audience was a reformulated version of the questionnaire for the defendants so that answering the questions from a third person perspective was possible (e.g., "How fair was the sentence you received?" reformulated in "How fair was the sentence the defendant received?").

*Additional measures.* The observers recorded an identification code for the judge who conducted the trial, the judges' sex, presence of a jury, presence of an attorney, whether the trial was conducted at a County or District court, whether the trial was an appellate proceeding or not, whether the trial was conducted in front of a juvenile court or not, the trials' length, the defendant's plea, whether the defendant used his right to remain silent, the number of the defendant's criminal records, the indicted crime and the judgment, as well as the sentence. It was assumed that the verdicts can be ordered according to their severity on an ordinal scale to capture the absolute result of the trial. This assumption might be criticized because a general evaluation of the severity of a sentence is hardly possible without taking the specifics of the case into account, especially if the sentences are of different qualities, like fines and prison sentences, for example (Gabriel & Greve, 2008). However, with this assumption we could include a measure for the outcome of the trial next to the relative ratings by the defendants. The present study assumed the following order of the verdicts, which starts with the least severe and ends with the most severe: Acquittal/ suspension of the proceedings, suspension of the proceedings on certain conditions imposed on the defendant, German juvenile law sentences warning, conditions and instructions, fines, detention of a juvenile delinquent, prison sentence on probation, court ordered sentence to a psychiatric or rehabilitation clinic, prison sentence.

## Results

### *Descriptive results*

In 9.2 percent of the cases the defendant was acquitted. In 11.4 percent the proceedings were suspended without imposing or on certain conditions imposed on the defendant. All the other cases resulted in a verdict of guilt. The most frequently imposed penalties were fines (27.6 percent) and sentences to prison on probation (21.8 percent). In 10.3 percent of the cases the defendants were sent to prison. The special German juvenile law sanctions warning, conditions, instructions and detention were imposed on 15.9 percent of the defendants. In 2.2 percent the defendants were

sent to a psychiatric or rehabilitation clinic. Fines ranged from 25 to 115 daily units of the defendants' assumed income ( $M = 50.87$ ,  $SD = 24.06$ ). Suspended prison sentences ranged between 3 and 24 month ( $M = 11.32$ ,  $SD = 6.40$ ) and prison sentences between 5 and 114 month ( $M = 29.44$ ,  $SD = 25.93$ ).

For the descriptive results of the defendants' ratings and the observed behavior of the judges see Tables 2, 3, and 4. Some variables of the coding scheme for the judges' behavior were characterized by large ceiling or bottom effects ( $M > 4.5$  or  $M < 1.5$  on a scale from 1 to 5;  $SD$ : .36 - .74; frequency of interruptions by the judge while the defendant was talking about emotions, perception of emotional meaning, the judge's compassion, the frequency of using special legal terminology, direct reference to the wording of the law or specific clauses, using the word please and using of commands). Therefore, all of these variables were excluded from further analysis.

Because of high correlations in the remaining set of coding scheme variables, a principal component analysis (PCA) was conducted to reduce the set of variables. For conducting the PCA the hierarchical data structure, with nesting of examinations under different judges, was ignored. For practicality reasons this procedure is acceptable according to Reise, Ventura, Neuchterlein, and Kim (2005) especially if no statistical inferences are drawn as it is the case with the PCA. Also, since standard literature on PCA includes discussions of the implementation of this method on data with multi-level structure (Rencher, 2003), ignoring of the hierarchical structure seemed warranted. With the data observed in  $n = 137$  defendants' examination, a PCA was conducted on the 12 remaining variables of the coding scheme. The data did not include any univariate or multivariate outliers. The Kaiser-Meyer-Olkin coefficient indicated that the correlations were of acceptable size for conducting a PCA on the data ( $KMO = .70$ ). Three components were retained, based on the eigenvalue  $> 1$  criterion. These three factors explained 64 percent of the total variance. The components had eigenvalues of 4.35, 1.97, and 1.38. After orthogonal rotation (varimax rotation) of the factor pattern matrix, the principal components were labeled as follows (see Table 5): Interested Listening, Respectful Treatment and Empathetic Listening. A variable was said to load on a particular factor if the loading was .40 or higher without having a loading of .40 or higher on another one.

The individual values characterizing a specific defendant's examination by a judge on the dimensions Interested Listening, Respectful Treatment and Empathetic Listening were computed by averaging the values of the coding scheme variables

listed in Table 5 instead of using factor scores. From a methodological perspective using factor scores would be preferable because they represent the case's value on that specific dimension. However, since the variables related to emotions were only considered if the observers rated addressing emotions as necessary, there were lots of missing values for cases where addressing emotions was rated as rather not necessary, so that factor scores could not be calculated in these cases. The empirical PCA aggregation of coding scheme variables matches the underlying assumption of its construction. The dimension Respectful Treatment encompasses all variables included in the coding scheme to measure the judges' respectful treatment. The variables that were included in the coding scheme to measure the trustworthiness of the judge, reflected in his/ her behavior, consist of two dimensions that are correlated to a low to medium size ( $r = .38, p < .01, n = 74$ ).<sup>5</sup>

The age of the judges as a proxy for the judges' tenure and experience was correlated with the judges' Respectful Treatment of the defendant ( $r = -.22, p < .05, n = 132$ ) in that younger judges treated the defendants more respectfully. The judges Interested Listening to the defendant was also negatively correlated to the judges' age ( $r = -.35, p < .01, n = 128$ ). The younger the judge, the more interested he/ she listened to the defendants. The judges' empathetic listening was not correlated with the judges' age.

### *Main analyses*

To test the competing research questions if a) the defendants' procedural justice judgments are based on the actual, objective treatment in the courtroom or b) the previously established relationship between the trial characteristics and the defendants' justice perception was largely influenced by a same source bias, we performed a number of correlation analyses. With these analyses we examined the relationship between ratings stemming from different data sources. First, we analyzed the correlation between the defendants' and the audiences' procedural justice perception as well as the relationship between the ratings of different procedural aspects from the two data sources defendants and audience. Second, we analyzed the relationship between the defendants' and the audiences' procedural justice ratings and the observed behavior of the judges.

Some of the variables were characterized by rather skewed distributions with a normal distribution of these variables being questionable that could not be normalized to a satisfying degree with different forms of transformation (see descriptive statistics

in Tables 2, 3, and 4). Therefore, we computed both the Pearson and the Spearman rank correlation ( $r_s$ ) and present both if inferences based on the two measures differed.

Before conducting the main correlational analyses the data was inspected for outliers. Outliers were defined as cases with standardized  $z$ -values larger than 3.29 or smaller than -3.29 that appeared disconnected from the other cases (Tabachnick & Fidell, 2007). In total, there were ten univariate outliers on all ratings by the defendants and the audience. As recommended by Tabachnick and Fidell (2007), variable values in these cases were replaced by a value that was half a scale point higher than the rating in the next extreme case.

Prior to addressing the main research question, the validity of the procedural justice measure was checked by performing a correlational analysis between the defendants' ratings of key procedural aspects and their procedural justice perception as performed in previous research. The highest correlations with the defendants' procedural justice rating existed with perceived fairness of procedural rules and the adequacy of the courts' information as basis for a judgment (see Table 4). The correlations are in line with earlier research. These results speak for the validity of the overall measure of the defendants' procedural justice impression.

In 79 trials data from the audience as a data source was gained. In 42.65 percent of these cases more than one person from the audience was willing to participate in the present study ( $M = 2.07$ ,  $SD = 1.74$ ). Before including the measures gained from this data source in the main analysis, the inner-audience agreement for the ratings of key procedural aspects from people present at the same trial was assessed. Therefore, we computed the intra-class correlation ( $ICC$ ; Hox, 2002). The  $ICCs$  ranged between .04 (equality of weighting the evidence) and .42 (outcome relative to previous expectations;  $M = .22$ ,  $SD = .11$ ). The rating of equality of weighting the evidence was characterized by a low inner audience agreement and therefore was not considered in further analysis. For all the other variables there was a moderate or good inner-audience agreement (range:  $ICC = .09 - .43$ ). Therefore, the different ratings from the audience collected in one trial were averaged if multiple persons agreed to participate in the present study.

*Data source defendants and audiences.* The correlation between the defendants' and the audiences' procedural justice perception was of medium size ( $r = .27$ , n.s.;  $r_s = .34$ ,  $p < .05$ ). Next to procedural justice, the defendants as well as the

audience rated their subjective impression of key procedural aspects of the trial. The correlations between the ratings gathered from the different data sources defendant versus audience were of small (negative) size or they were close to zero (see Table 6).

*Data source defendants and observers.* The same pattern arose when looking at the correlations between the defendants' procedural justice rating of a trial and the observed behavior of the judge during that trial. The correlation between the ratings, gained from the data source defendants, and judges' behavior coded by the observers, as an independent source of data, were small and insignificant (Interested Listening:  $r = .08$ , Respectful Treatment:  $r = -.15$ , Empathetic Listening:  $r = -.03$ ). For the global rating of the judges, the defendant and the observers partly used the same items. The agreement between the defendants and the observers on these items was low (performance of duties:  $r = .22$ ,  $p < .01$ ,  $r_s = .14$ , n.s.,  $n = 86$ ; courtesy:  $r = .18$ , n.s.,  $n = 84$ ; sympathy:  $r = .36$ ,  $p > .01$ ,  $r_s = .16$ , n.s.,  $n = 85$ ).

*Data source audience and observers.* The pattern was different when data gathered from the audience were considered. Here, there existed a relation between the judges' behavior measured via observation and the audience's procedural justice rating. The more respectful the judge treated the defendant, the fairer the trial was perceived by the audience ( $r = .42$ ,  $p < .01$ ). There was no relation between the judges' Interested and Empathetic Listening measured via observation and the audience's procedural justice perception. ( $r = .03$ , n.s.;  $r = .08$ , n.s.). To judge the relative importance of the relation between observed respectful behavior and the audiences' procedural justice rating, especially in comparison to the subjective assessment of procedural aspects by the audience, a prediction model was estimated. The data of the present research were characterized by a hierarchical structure with observations and ratings, gained in multiple trials, nested within different judges. Despite this structure, we did not conduct a multi-level model (Hox, 2002; Luke, 2004) because the sample size of the present research was rather small for achieving a stable estimate and because of a facilitated interpretation of a multiple regression analysis without accounting for the hierarchical data structure. To capture all the dimensions characterizing fair procedures identified in previous research the following predictors were selected: The audiences' rating of the defendant's opportunity to present evidence, the adequacy of the court's information as basis for a decision, a general evaluation of the judge and the trial's outcome relative to expectations. Additional to these subjective ratings the respectfulness of the judges'

treatment, measured via observation, was included as a predictor to test the predictive power of this more objective measure in comparison to the audiences' subjective ratings of key procedural aspects. For theoretical reasons and for most parsimonious model building the interactions between these variables were not included in the model. All data collected from the audience in  $n = 77$  trials were analyzed. Inspection of the residual plots suggested that the statistical assumptions for the analysis were met. The model accounted for a substantial percentage of the variance ( $R^2 = .55$ ,  $R^2_{adj} = .52$ ). In addition to the effects of the subjective measures on the procedural justice rating (see Table 7), there was a significant effect of the respectfulness of the judge's treatment, measured via observation. The more respectful the judge treated the defendant, the fairer the trial was perceived by the audience ( $b = .27$ ,  $p < .05$ ).

### Discussion

Previous research established a relationship between the perception of key procedural aspects and of procedural justice. The more the defendants felt being actively involved in the decision making process, with an opportunity to state their own view, as well as the more they felt being treated with respect and dignity, the fairer a procedure was perceived. Additionally, previous empirical findings suggest that the more the defendants perceived the judge as honest, objective and impartial as well as compassionate and interested, the fairer the procedure was perceived (Tyler, 2000). The present study replicated these findings: the perception of the treatment in court was related to the fairness perception, in that ratings of trial characteristics like the opportunity to present evidence, the appropriateness of the court's information as the basis for a judgment and the time the court took to consider the case were highly correlated with the procedural justice perception.

The main focus of the present research was to address a shortcoming of previous justice research that mostly used only one data source (e.g., subjective ratings of the defendants). As demanded in the procedural justice literature (e.g., Colquitt et al., 2001; MacCoun, 2005), the present research used a multi-source, multi-method approach to analyze the relationship between informal aspects of the treatment in court by measuring the judges' behavior and perceived procedural justice as well as the differences in the justice perception with data from different sources (defendants, audience). If the procedural justice judgments by the defendant are based on the actual, objective treatment in court, and thus on objective characteristics of the

trial, then there should be a relation between the judges' behavior measured via observation and the defendants' procedural justice ratings. However, if the above mentioned relationship between trial characteristics and the defendants' justice impression mainly resulted from using the same data source for these measures, the relationship between the judges' behavior and the defendants' justice ratings should be small. Also, there should be differences between the data stemming from different sources (subjective ratings by the defendants and by the audience). The results of the present research gained support for the latter proposition in that we found no correlations between the defendants' procedural justice ratings and the informal treatment in court by the judge, measured via observation of the judges behavior, but a significant positive correlation between the audience's procedural justice ratings and the observed respectful behavior of the judge. The more respectful the judge treated the defendant, the fairer the trial was perceived by the audience. Thus, for the audience we found evidence for the previously established connection between the informal quality of treatment and justice perception (Blader & Tyler, 2003). However, we did not find this relation for the defendants. Additionally, the correlations between the defendants' and the audiences' subjective ratings of procedural justice were of a medium size and the correlation between the defendants' and the audiences' rating of key procedural aspects were small or close to zero. Thus, we only found small relations between the perceptions of these two sources. One general reason for these differences in the defendants' and the audiences' perception of key procedural aspects and procedural justice, demonstrated with the present research, might be a difference in psychological distance. The psychological distance to the exercised criminal trial is smaller for the defendants that are directly involved in the trial and are directly affected by its outcome in comparison to the audience. This can lead to differences in cognition and motivation and, as a result, to differences in the perception of procedural justice. For example, Greifeneder, Müller, Stahlberg, Van den Bos, and Bless (2011) recently demonstrated that the ease or difficulty with which information relevant for justice judgments can be accessed from memory can influence procedural justice judgments. The easier it is to retrieve information related to unfair procedural aspects, the less positive a procedure was rated. With differences in distance the perceptual focus of people in the audience and the defendants is different. According to Construal Level Theory (Liberman & Trope, 1998, Liberman, Trope, & Stephan, 2007; Trope & Liberman, 2003) the psychological distance affects

the perception of an event in that people with higher distance perceive an event as more decontextualized, abstract and focus on its' essence instead of its' secondary features. With justice being an abstract value people in the audience might have kept this bigger-picture value more in mind while observing the trial and then, as a consequence, this might have affected which aspects were easy or difficult to retrieve from memory while making a justice judgment. Thus, psychological distance might have affected the defendants' and the audiences' procedural justice judgments. Also due to differences in psychological distance, the defendants' and the audiences' procedural justice judgments should be affected differently by self-serving biases. Francis-Gladney, Mager, and Welker (2010) demonstrated that receiving a favorable outcome led to increased procedural fairness perceptions because of self-serving attributions. People saw themselves as more responsible for outcomes that were the result of fair procedures. This self-serving bias should be more relevant for the defendants' justice perceptions because they were directly involved in the decision making and thus should affect the procedural justice judgments of the defendants more than the ones given by the audience. The audience's perception might thus be less biased and this might account for a stronger relation with the objective observational data. In addition to the influence of outcome favorability, procedural justice ratings are influenced by moral positions (Skitka & Houston, 2002). Regardless of how an outcome is achieved, people's fairness perceptions are influenced by their moral cognition. A trial is perceived as fairer when a person perceived to be guilty is convicted or when a person perceived to be innocent is acquitted in comparison to the opposite outcomes. Since the defendant's own sense of guilt and innocence might be different in comparison to the audience's impression this process should influence the justice ratings from these two data sources in different ways.

### *Limitations*

The latter two mentioned reasons for differences between the data from the defendants and the audience are also related to one major shortcoming of the present research. The justice ratings and the ratings of key procedural aspects by the audience and the defendants were only assessed after the verdict. In previous research, though, it has been established that these hind-sight fairness judgments are influenced by the favorability of the outcome (Francis-Gladney et al., 2010; Skitka, Winkvist, & Hutchinson et al., 2003). More favorable outcomes result in an increase of the

procedural fairness perception. Future research should improve this shortcoming within a multi-source approach and evaluate the differences in the fairness perception between the different data sources on all steps of a trial.

Another shortcoming of the present research is that in observation as a method for data collection generally suffers from a potentially low reliability (Amelang et al., 2006). However, it seems rather unlikely that the null finding for a relation between observed judges' behavior and defendants' procedural justice perception can be explained just by invalid observations. Considering that there was evidence for a positive correlation between the observed respectfulness of the judges' treatment and the audiences' procedural justice perception in the direction predicted by previous research and theory (Tyler, 2000), the null finding might rather result from the defendants not basing their justice impressions on the actual treatment in court, at least in a lesser extent than the audience. People from the audience, who were not directly involved in the trial, might just have had a more objective perception of the actual treatment.

### *Conclusion*

In conclusion, the present research highlighted the differences in the justice perception between different data sources. The results of the present study strongly suggest taking the data source into account while analyzing the relation between trial characteristics and procedural justice perception. Especially, if the ultimate goal is to infer aspects of trials that could be changed to improve the justice impressions, different sources of data should be taken into account.

## References

- Adams, J. S. (1965). Inequity on social exchange. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 267-299). New York: Academic Press.
- Amelang, M., Schmidt-Atzert, L., Fydrich, T., & Moosbrugger, H. (2006). *Psychologische Diagnostik und Intervention* (4<sup>th</sup> completely expanded and revised edition). Heidelberg: Springer Medizin.
- Aronsson, K., Jönsson, L., & Linell, P. (1987). The courtroom hearing as a middle ground: Speech accommodation by lawyers and defendants. *Journal of language and social psychology*, 6, 99-115.
- Bierbrauer, G., & Klinger, E. (2008). Verfahrensgerechtigkeit. In R. Volbert & M. Steller (Eds.), *Handbuch der Rechtspsychologie* (pp. 507-518). Göttingen: Hogrefe.
- Bierhoff, H. W. (2006). *Sozialpsychologie: Ein Lehrbuch* (6<sup>th</sup> revised and expanded edition). Stuttgart: Kohlhammer.
- Bies, R. J., & Moag, J. F. (1986). Interactional justice: Communication criteria of fairness. In R. J. Lewicki, B. H. Sheppard, & M. H. Bazerman (Eds.), *Research in negotiations in organizations* (Vol. 1, pp. 43-55). Greenwich, CT: JAI Press
- Bintig, A. (2004). „Empathie“ in Täter-Opfer-Verhältnissen – Konsequenzen für die psychotherapeutische Praxis. *Gruppendynamik und Organisationsberatung*, 35, 67-82.
- Blader, S. L., & Tyler, T. R. (2003). A four-component model of procedural justice: Defining the meaning of a ‚fair‘ process. *Personality and Social Psychology Bulletin*, 29, 747-758.
- Colquitt, J. A. (2001). On the dimensionality of organizational justice: A construct validation of a measure. *Journal of Applied Psychology*, 86, 386-400.
- Colquitt, J. A., Conlon, D. E., Wesson, M. J., Porter, O. L. H., & Ng, K. Y. (2001): Justice at the Millennium: A meta-analytic review of 25 years of organizational justice research. *Journal of Applied Psychology*, 86, 425-445.
- 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, 137-150.

- Eilsberger, R. (1969). Die Hauptverhandlung aus der Sicht jugendlicher und heranwachsender Angeklagter. *Monatsschrift für Kriminologie und Strafrechtsreform*, 52, 304-313.
- Francis-Gladney, L., Magner, N. R., & Welker, R. B. (2010). Does outcome favorability affect procedural fairness as a result of self-serving attributions? *Journal of Applied Social Psychology*, 40, 182-194.
- Gabriel, U., & Greve, W. (2008). Strafbedürfnisse und Strafeinstellungen. In R. Volbert & M. Steller (Eds.), *Handbuch der Rechtspsychologie* (pp. 467-476). Göttingen: Hogrefe.
- Greenberg, J. (1993). The social side of fairness: Interpersonal and informational classes of justice. In R. Cropanzano (Ed.), *Justice at the workplace: Approaching fairness in human resource management* (pp. 79-103). Hillsdale, NJ: Erlbaum.
- Greifeneder, R., Müller, P., Stahlberg, D., Van den Bos, K., & Bless, H. (2011). Beyond procedure's content. The role of accessibility experiences and personal uncertainty in procedural justice judgments. *Experimental Psychology*, 58, 341-352.
- Haller, V., & Machura, S. (1995). Procedural justice at German courts as seen by defendants and juvenile prisoners. *Social Justice Research*, 8, 197-215.
- Haller, V., Machura, S., & Bierhoff, H. W. (1995). Urteil und Verfahrensgerechtigkeit aus der Perspektive jugendlicher Strafgefangener. In G. Bierbrauer, W. Gottwald & B. Biernbreier-Stahlberger (Eds.), *Verfahrensgerechtigkeit. Rechtspsychologische Forschungsbeiträge für die Justizpraxis* (pp. 111-137). Köln: Dr. Otto Schmidt.
- Hox, J. (2002). *Multilevel analysis. Techniques and applications*. Mahwah: Lawrence Erlbaum.
- Leventhal, G. S. (1980). What should be done with equity theory? In K. J. Gergen, M. S. Greenberg, & R. H. Willis (Eds.) *Social exchange: Advances in theory and research*. New York: Plenum.
- Liberman, N., & Trope, Y. (1998). The role of feasibility and desirability considerations in near and distant future decisions: A test of temporal construal theory. *Journal of Personality and Social Psychology*, 75, 5-18.

- Liberman, N., Trope, Y., Stephans, E. (2007). Psychological distance. In A. W. Kruglanski & E. T. Higgins (Eds.), *Social Psychology: Handbook of basic principles* (2<sup>nd</sup> ed., Vol. 2, pp. 353-383). New York, NY: Guilford Press.
- Lind, E. A. (1995). Verfahrensgerechtigkeit und Akzeptanz rechtlicher Autorität. In G. Bierbrauer, W. Gottwald & B. Birnbreier-Stahlberger (Eds.). *Verfahrensgerechtigkeit. Rechtspsychologische Forschungsbeiträge für die Justizpraxis* (pp. 3-19). Köln: Dr. Otto Schmidt.
- Lind, E. A., & Tyler, T. R. (1988). *The social psychology of procedural justice*. New York: Plenum Press.
- Luke, D. A. (2004). *Multilevel modeling*. Thousand Oaks: Sage.
- MacCoun, R. J. (2005). Voice, Control, and Belonging: The double-edged sword of procedural fairness. *Annual Review of Law and Social Science*, 1, 171-201.
- Mischo, C. (2003). Wie valide sind Selbsteinschätzungen der Empathie? *Gruppendynamik und Organisationsberatung*, 34, 187-203.
- Montada, L., & Karls, E. (2007). *Mediation. Ein Lehrbuch auf psychologischer Grundlage* (2<sup>nd</sup> completely expanded and revised edition). Weinheim: Beltz.
- Reise, S. P., Ventura, J., Neuchterlein, K. H., & Kim, K. H. (2005). An illustration of multilevel factor analysis. *Statistical development and applications*, 84, 126-136.
- Rencher, A. C. (2002). *Methods of multivariate analysis*. New York: Wiley.
- Rohrmann, B. (1978). Empirische Studie zur Entwicklung von Antwortskalen für sozialwissenschaftliche Forschung. *Zeitschrift für Sozialpsychologie*, 9, 222-245.
- Rosenthal, R. (1987). *Judgment studies. Design, analysis, and meta-analysis*. Cambridge: Cambridge University Press.
- Statistisches Bundesamt (2009). *Fachserie 10. Reihe 3. Rechtspflege. Strafverfolgung. 2007*. Wiesbaden: Statistisches Bundesamt.
- Skitka, L. J., & Houston, D. A. (2002). When due process is of no consequences: moral mandates and presumed defendant guilt or innocence. *Social Justice Research*, 14, 305-326.
- Skitka, L. J., Winquist, J., Hutchinson, S. (2003). Are outcome fairness and outcome favorability distinguishable psychological constructs? A meta-analytical review. *Social Justice Research*, 16, 309-341.

- Streicher, B., Jonas, E., Maier, G. W., Frey, D., Woschee, R., & Wassmer, B. (2008). Test of the construct and criteria validity of a German measure of organizational justice. *European Journal of Psychological Assessment, 24*, 131-139.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5<sup>th</sup> international edition). Boston: Pearson Allyn and Bacon.
- Tausch, A. M., & Langer, I. (1971). Soziales Verhalten von Richtern gegenüber Angeklagten; Merkmale, Auswirkungen sowie Änderungen durch ein Selbst-Training. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 4*, 283-303.
- Thibaut, J., & Walker, L. (1975). *Procedural justice: A psychological analysis*. Hillsdale: Lawrence Erlbaum.
- Trope, Y., & Liberman, N. (2003). Temporal construal. *Psychological Review, 110*, 403-421.
- Tyler, T. R. (1984). The role of perceived injustice in defendants' evaluation of their courtroom experience. *Law and Society Review, 18*, 51-74.
- Tyler, T. R. (1989). The psychology of procedural justice: A test of group value model. *Journal of personality and social psychology, 55*, 333-344.
- Tyler, T. R. (2000). Social justice: outcome and procedure. *International Journal of Psychology, 35*, 117-125.
- Tyler, T. R. & Lind, E. A. (1992). A relational model of authority in groups. In M. P. Zanna (Ed.). *Advances in experimental social psychology*, Vol. 25. (pp. 115-191). San Diego: Academic Press.
- Wolff, S. & Müller, H. (1997). *Kompetente Skepsis. Eine Konversationsanalytische Untersuchung zur Glaubwürdigkeit in Strafverfahren*. Wiesbaden: Westdeutscher Verlag.

## Footnotes

<sup>1</sup> Cases with multiple defendants were treated as separate cases in the statistical analysis. Even though the hierarchical nesting of two or more defendants at one trial should be accounted for in the statistical analysis, such an analysis was not possible due to the small amount of cases with multiple defendants. The potential model violation resulting from disregarding this nesting is due to correlated errors in the absolute residuum of the model. This correlation would reflect a higher similarity of defendants indicted at the same trial. Given the small number of these cases, treating them as separate cases in the statistical analysis seems warranted considering that potential model violations can only be minor with an the advantage of an increased sample size.

<sup>2</sup> The following statements refer to the data collected in cases in which the defendant was willing to participate in the study if not stated otherwise.

<sup>3</sup> The length of the trial was negatively correlated with the procedural justice perception of the defendants in that longer trial were perceived as less fair ( $r = -.22$ ,  $p < .05$ ). This correlation might be driven by the fact that longer trials were more likely to result in a more severe sentence and thus a less favorable outcome reflected in a positive Spearman rang correlation between absolute outcome and trial length ( $r_s = .45$ ,  $p < .001$ ). The length of the trial had no incremental effect in a regression including predictors for the perception of the outcome and trial characteristics on the perceived procedural justice of the defendants. The trial length was uncorrelated with the audience's procedural justice rating ( $r = -.03$ ).

<sup>4</sup> There was no effect of knowing the defendant before the trial on the audiences' justice perception. Therefore, this variable is not discussed in the following sections.

<sup>5</sup> The judges' age was uncorrelated with the defendants' as well at the audiences' procedural justice perception.

## Table Captions

*Table 1.* Sample of crimes for cases were defendants were willing to participate in the study in comparison to the official statistics on condemned persons in Germany

\**Note:* Freq = absolute frequency of cases  
% = percental frequency of cases

*Table 2.* Descriptive statistics and inter-rater reliability of coding scheme variables

*Table 3.* Descriptive statistics and inter-rater reliability of the global rating of the judge

*Table 4.* Descriptive statistics and Correlational Analysis

\**Note:* \* Correlation significant at  $p < .05$  \*\* correlation significant at  $p < .01$   
<sup>a</sup> all correlations of the variable absolute outcome with other variables are Spearman correlations

*Table 5.* Principal component analysis of the twelve coding scheme variables. Three-factor solution. The bold numbers indicate loadings of coding scheme variables that were included in each particular component

\**Note:* Listwise (loadings  $< 0.4$  were removed for most clarity)  
<sup>a</sup> different sign because of the different scale polarity

*Table 6.* Correlations between the ratings made by the defendants and the audience

\**Note:* If inferences based on Pearson and Spearman correlations were different, both coefficients are listed (Spearman Correlation behind the dash)  
\* Correlation significant at  $p < 0.05$  \*\* correlation significant at  $p < 0.01$

*Table 7.* Coefficients, standardized coefficients and semipartial correlation for predicting the audiences' perceived procedural justice rating

\**Note:* All variables were centered up on their mean value before conduction the analysis  
\* regression coefficient significant at  $p < .05$  \*\* regression coefficient significant at  $p < .01$

Table 1

	Sample of crimes		Official statistic	
	Freq	%	Freq	%
Criminal offense in total	87		863 239	
Crimes against the physical integrity	18	20.69	89 590	10.38
Other property crimes and forgery of documents	14	16.09	252 530	29.25
Crimes against narcotic acts	14	16.09	52 752	6.11
Theft, unlawful appropriation	11	12.64	146 602	16.98
Traffic offenses	10	11.49	218 146	25.27
Robbery and armed robbery	7	8.05	6 818	0.79
Offences against the person other than traffic offenses	5	5.75	49 897	5.78
Crimes against sexual self-determination	4	4.60	9 248	1.07
Others: Criminal damage, Committing offences in a senselessly drunken state, Misleading the authorities about the commission of an offence, violation of the weapon law	4	4.60		

Table 2

	Examination of the defendants			Inter-rater Reliability	
	n	M	SD	n	R <sub>Int</sub>
Defendant's opportunity to state his/ her view	133	4.29	.71	479	.62
Judge's tone	137	3.18	.73	483	.76
Frequency impression of interruptions by the judge	137	4.00	.87	483	.83
Frequency impression of interruptions by the judge while the defendant was talking about emotions	74	1.11	.37	483	.82
Frequency impression of disrespectful comments	137	3.61	.96	483	.82
Judge's compassion	74	1.44	.69	483	.87
Judge's empathy	74	1.93	.92	483	.88
Judge's perspective-taking	137	2.44	1.01	483	.83
Judge's interest	137	4.03	.70	483	.67
Frequency impression of questions on emotions	74	1.57	.74	483	.88
Frequency impression of perception of emotional meaning	74	1.55	.74	483	.77
Frequency impression of eye-contact	137	3.47	.70	483	.73
Frequency impression of expressions of comprehension	137	3.50	1.17	483	.81
Frequency impression of paraphrasing	137	2.76	1.13	483	.80
Frequency impression of the use of special legal terminology	137	4.75	.40	483	.84
Frequency impression of the use of complex language	137	4.87	.36	483	.74
Frequency impression of direct reference to the wording of the law or specific clauses	137	4.80	.48	483	.86
Frequency impression of the use of the word please	137	1.17	.40	483	.82
Frequency impression of the use of commands	137	4.66	.52	483	.73
Frequency impression of reversible expressions	137	3.87	.90	482	.84
Necessity to address emotions	136	2.64	1.27	483	.92

Table 3

	Examination of the defendants			Inter-rater Reliability	
	n	M	SD	n	R <sub>Int</sub>
Unappreciative/ appreciative	137	3.21	.72	129	.75
unfriendly/ friendly	137	3.56	.72	129	.77
cold-hearted/ warm-hearted	137	2.76	.74	129	.73
Impersonal/ personal	137	2.94	.94	129	.73
criticizing/ accepting	137	2.73	.79	129	.77
Relentless/ compliant	137	3.07	.84	129	.76
impatient/ patient	137	3.25	.91	129	.73
courtesy	137	3.80	1.55	129	.83
overall performance of duties	137	3.00	1.57	129	.81
honesty	137	3.10	.85	129	.55
sympathy	137	3.46	1.57	129	.86

Table 4

	n	M	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Outcome level																	
1 absolute	123																
2 Relative to expectations	87	5.43	2.53	-.26*													
3 Relative to others	81	5.52	2.30	-.21	.52**												
4 Satisfaction with outcome	86	5.48	2.79	-.38**	.78**	.52**											
Fairness																	
5 Distributive justice	87	5.49	2.64	-.25*	.71**	.56**	.83**										
6 Procedural justice	87	5.48	2.39	-.31**	.60**	.50**	.75**	.70**									
Procedural aspects																	
7 opportunity to present evidence	86	5.93	2.40	-.23*	.45**	.48**	.64**	.56**	.55**								
8 adequacy of the court's information	86	6.14	2.52	-.07	.50**	.36**	.54**	.55**	.55**	.38**							
9 time the court took to consider the case	85	6.26	2.11	-.16	.46**	.40**	.58**	.58**	.54**	.50**	.49**						
10 impartiality of the court	84	3.62	2.90	.12	-.13	-.11	-.11	-.05	-.08	-.01	-.05	-.19					
11 taking evidence into account	85	5.72	2.23	-.11	.45**	.21	.46**	.55**	.52**	.32**	.53**	.40**	-.09				
12 equality of weighting evidence	85	5.28	2.38	-.13	.47**	.36**	.54**	.59**	.44**	.33**	.34**	.44**	-.07	.40**			
13 fairness of procedural rules	87	5.24	2.27	-.27*	.58**	.44**	.69**	.55**	.75**	.57**	.49**	.59**	.02	.41**	.47**		
14 Evaluation of the judge	83	6.57	1.79	-.38**	.58**	.49**	.73**	.59**	.54**	.53**	.41**	.50**	-.13	.28*	.43**	.57**	
15 Evaluation of German courts	84	4.62	2.29	-.03	.32**	.24*	.50**	.47**	.43**	.32**	.20	.35**	.08	.26*	.53**	.50**	.29**

Table 5

	Factor		
	Interested Listening	Respectful Contact	Empathetic Listening
Judge's interest	<b>.86</b>		
Defendants' opportunity to present his/ her view	<b>.71</b>		
Frequency impression of expressions of comprehension	<b>.71</b>		
Frequency impression of paraphrasing	<b>.63</b>		
Frequency impression of disrespectful comments		<b>.90</b>	
Frequency impression of reversible expressions		<b>.88</b>	
Judge's tone		<b>.83</b>	
Frequency impression of interruptions by the judge <sup>a</sup>	-.49	.53	
Frequency impression of questions on emotions			<b>.75</b>
Judge's empathy			<b>.72</b>
Judge's perspective-taking	.51		.66
Frequency impression of eye-contact			<b>.52</b>

Table 6

	n	r
Outcome		
Outcome relative to expectations	44	.45**
Outcome relative to others	39	.24
Outcome satisfaction	44	.40**
Fairness		
Distributive justice	44	.34*
Procedural justice	43	.27/ .34*
Procedural aspects		
opportunity to present evidence	44	.35* / .28
adequacy of the court's information	44	.10
time the court took to consider the case	44	-.06
impartiality of the court		.07
taking evidence into account	43	-.18
equality of weighting evidence		.11
Fairness of procedural rules	44	.20
Evaluation of the judge	42	.55**
Evaluation of German courts	42	-.04

Table 7

	B	SE B	$\beta$	sr <sup>2</sup>
constant	6.58**	.09		
opportunity to present evidence	.52**	.12	.35	.29
adequacy of the court's information	.11 (p=.09)	.06	.13	.12
evaluation of the judge	.47**	.11	.38	.29
outcome relative to expectations	.07	.05	.10	.10
respectful treatment	.27*	.12	.18	.16