Is it possible to improve test takers’ perceptions of ability tests by providing an explanation?

Klaus G. Melchers
Universität Ulm, Germany

and

Barbara Körner
Universität Zürich, Switzerland

Author Note
We would like to thank Kim Bentlage, Lisa Leusch, Julia Mies, Linda Miller, Anja Roth, and Ebru Sezer for their help with data collection.

Address for correspondence:
Klaus Melchers
Institut für Psychologie und Pädagogik
Abteilung Arbeits- und Organisationspsychologie
Universität Ulm
Albert-Einstein-Allee 41
89069 Ulm
Germany
Phone: +49-731-50 31740
Fax: +49-731-50 31749
Email: klaus.melchers@uni-ulm.de
Abstract

Previous meta-analytic findings revealed that explanations can improve applicants’ perceptions of selection procedures. However, they also suggest that these positive effects do not generalize to ability tests. However, given some limitations of previous studies and the small empirical basis for the corresponding meta-analytic results, we had another look at whether perceptions of ability tests can be improved by providing an explanation. In two experimental studies, participants had to complete either an attention or a general mental ability test. In the explanation group, a justification was given concerning the content, relevance, and predictiveness of the test. In contrast, no explanation was given in the control group. Providing an explanation significantly improved test takers’ fairness perceptions.

Keywords: applicant reactions; explanations; procedural fairness; ability tests
Is it possible to improve test takers’ perceptions of ability tests by providing an explanation?

In addition to the psychometric properties of a selection procedure, it is also important to consider how a procedure is perceived by applicants because these perceptions are related to several relevant outcomes including perceptions of organizational attractiveness, intentions to proceed with a selection process, to accept a job offer, to reapply, or to recommend the organization to others (cf. Hausknecht, Day, & Thomas, 2004; or McCarthy et al., 2017, for reviews). Furthermore, recent research also found effects of applicants' perceptions on actual job offer acceptance (Harold, Holtz, Griepentrog, Brewer, & Marsh, 2016; Konradt, Garbers, Böge, Erdogan, & Bauer, 2017). In light of these findings, it seems unfortunate that valid psychometric tests are often perceived less positively in comparison to other more expensive selection procedures like work samples or interviews (Anderson, Salgado, & Hülsheger, 2010; Hausknecht et al., 2004).

Providing explanations to applicants has been suggested as a cost effective and straightforward way to positively influence their perceptions of such tests and meta-analytic work by Truxillo, Bodner, Bertolino, Bauer, and Yonce (2009) found that explanations can have beneficial effects on perceptions of a selection procedure. Unfortunately, this meta-analysis did not confirm such a beneficial effect of explanations for ability tests such as tests of general mental ability (GMA) or of specific abilities. This is rather unfortunate given the good validity of ability tests (e.g., Bertua, Anderson, & Salgado, 2005; Schmidt & Hunter, 2004). However, a limitation of Truxillo et al.’s meta-analytic result is that it is based on a rather small empirical basis. Furthermore, many of the previous studies on effects of explanations on perceptions of ability tests did not use explanations that were related to the content and job relevance of the test.

The aim of our research was to address these limitations and thereby to extend previous research on the beneficial effects of explanations. Specifically, we hypothesized that providing a justification that focuses on the content of a test and the relevance and
predictiveness of it can improve test takers’ perceptions of this test – similar to findings for other selection procedures (Truxillo et al., 2009). Accordingly, we set up two studies to test whether this is indeed the case.

**Review of Previous Research**

During the last decades, several theoretical frameworks have been put forth that describe relevant determinants of applicant reactions (cf. Truxillo & Bauer, 2011). The most influential framework was proposed by Gilliland (1993; also see Gilliland & Steiner, 2012) whose model of applicant reactions is based on organizational justice theory. It defines a set of justice rules that are pertinent to the perceived fairness of a selection process and that comprise aspects of procedural, distributive, informational, and interpersonal justice. According to the model, these justice rules determine the perceived fairness of a selection procedure and applicants react negatively in situations when the rules are violated. Most of the applicant reactions research during the last two decades used this model (see, for example, reviews by McCarthy, Bauer, Truxillo, Anderson, et al., 2017; or by Truxillo & Bauer, 2011) and most of Gilliland’s rules were shown to be relevant for subsequent applicant reactions (Hausknecht et al., 2004; McCarthy, Bauer, Truxillo, Anderson, et al., 2017). Furthermore, in their recent review of applicant reactions research, McCarthy, Bauer, Truxillo, Anderson, et al. noted that the provision of explanations to applicants was one of the aspects that were most consistently related to fairness perceptions.

Conceptually, explanations can contribute to several of the procedural justice rules suggested in Gilliland’s (1993) model. First, explanations are relevant with regard to the provision of information about the administration of selection procedures or the necessity to use certain procedures. Furthermore, explanations can also be used to provide information on the job relatedness and predictiveness of a selection procedure. And explanations can also be used to provide information about other applicants (e.g., in case of a rejection that others were
more qualified) or about practices by other organizations that use or do not use similar selection procedures.

Possible explanations fall into two major categories: Excuses and justifications. Excuses are explanations in which the actor denies full responsibility for a process or an outcome because this process or outcome was (at least to some degree) beyond his or her control. In contrast to this, justifications are explanations in which the actor stresses the appropriateness or ethicality of a process or outcome. According to Truxillo et al.’s (2009) meta-analysis, both justifications and excuses have positive effects on applicants’ perceptions, but especially justifications seem to be beneficial for perceptions of a selection procedure. Furthermore, explanations also contributed to improved perceptions of the organization and to an enhancement of applicants’ test taking motivation. Finally, according to the meta-analysis, the positive effects of explanations hold regardless of the selection outcome.

Unfortunately, as already noted above, Truxillo et al. (2009) found no beneficial effect of explanations that were related to ability tests such as GMA tests or tests of specific abilities. And even though previous research found that applicants in general perceive GMA tests as more favorable in comparison to some other selection procedures such as personality or integrity tests (Anderson et al., 2010; Hausknecht et al., 2004), this relatively positive perception is not true for all ability tests. A comparison of several common German ability tests by Kersting (2008), for example, revealed considerable differences with regard to the overall acceptability of different tests and found that especially more abstract tests were perceived rather negatively by test takers. Accordingly, it would be beneficial if these perceptions could be improved by explanations.

A limitation of Truxillo et al.’s (2009) meta-analytic result for ability tests is that it rests on a rather small empirical basis. Specifically, the meta-analytic estimate for ability tests is based on only six primary studies. And we are not aware of any additional studies since then that dealt with the effects of explanations on test takers’ perceptions of ability tests even
though Hausknecht et al. (2004) already called for more research “to examine the potential benefits of interventions that could improve applicant reactions … such as providing explanations” (p. 674) quite some time ago.

With regard to specific primary studies that investigated effects of explanations in the context of ability tests – and in line with Truxillo et al.’s (2009) meta-analytic result – positive effects were only found in a few cases. Specifically, LaHuis, Perreault, and Ferguson (2003) found higher procedural fairness perceptions and more positive perceptions of content and predictive validity after an explanation and Rolland and Steiner (2007) found positive effects on perceived predictive validity and propriety of questions. Both studies used selection simulations and participants were told that the respective tests were good predictors of job performance. In addition, Rolland and Steiner also provided a justification for the alleged selection decision in which participants were given “comparative information of participants’ scores to the results of other ‘applicants’” (p. 2808). However, in both studies participants had to complete not only a GMA test but also a personality test (LaHuis et al., 2003) or an overt integrity test (Rolland & Steiner, 2007) and the explanation was directed at both tests at the same time. Furthermore, specific effects concerning perceptions of the GMA test were not collected in the former study and not reported in the latter one.

In contrast to these few positive effects of explanations on perceptions of ability tests, results from other studies (Horvath, Ryan, & Stierwalt, 2000; Lievens, De Corte, & Brysse, 2003) or with additional groups or additional variables in the studies by LaHuis et al. (2003) and by Rolland and Steiner (2007) found non-significant effects when treatment groups were compared to a no-explanation control group. In these studies, participant were provided with information about test reliability and validity (Lievens et al., 2003), the use of much more rigorous tests by other organizations (Horvath et al., 2000), or the need for “a better way of weeding out poor performers” after a “wave of poor hiring decisions” (Horvath et al., 2000, p. 315). Furthermore, a more extensive explanation in LaHuis et al.’s (2003) study that also
covered information about the way how job requirements were identified and about test content (i.e., cognitive ability and personality) had no effect. Finally, an explanation that informed test takers about the hope that the test did not discriminate with regard to race, gender, age, and ethnicity even led to significantly lower face validity perceptions in comparison to the control group (Horvath et al., 2000).

As can be seen in this review of earlier research, many of the explanations that were used in previous studies with ability tests were not specifically related to the content and the relevance of the test for the job in question. Instead, they targeted aspects such as reduced discrimination or practices by other organizations or they contained relatively abstract information on reliability and validity. In contrast to this, research related to other selection procedures such as personality tests repeatedly found that explanations concerning the job relevance and the predictiveness of the test results led to improved perceptions of job-relatedness (e.g., Holtz, Ployhart, & Dominguez, 2005).

Given this state of previous research on effects of explanations on test takers’ perceptions of ability tests, we wanted to evaluate the hypothesis that a justification that focuses on test content and relevance as well as on the predictiveness of the test can improve test takers’ perceptions. To test this hypothesis, we conducted two studies. In Study 1, we used a specific ability test that was perceived rather negatively in previous research, and in Study 2, we used a GMA test to evaluate the generalizability of the results from the first study.

**Study 1**

**Method**

**Sample.** The sample consisted of 96 university students (mean age = 23.17 years, SD = 3.82; 49% females) most of whom were majoring in the natural sciences, engineering, or medicine. No psychology majors beyond the first semester were allowed to participate to
prevent that study participants had previous training in psychometric testing. Of the participants, 73% had previously already applied for a job and 28% had completed selection tests before. Additionally, 49% of the participants already had work experience.

**Procedure.** We asked participants to describe an attractive job in their field of study for which they would apply after completing their degree. Then, they were told that the following test is often used in selection settings and that successful completion of it influences hiring decisions. To further enhance their motivation for the later administration of the test (a common attention test, see below for more information), we promised participants gift certificates for a local cinema for the ten individuals with the highest test scores.

Following this instruction, participants were randomly allocated to the control group and the explanation group. Participants in the explanation group were provided with a justification containing three important pieces of information about the use of the attention test in selection contexts: Specifically, test takers were informed (a) that the test is an established procedure to measure attention and concentration, (b) that attention and concentration are important for challenging jobs that require simultaneous attention to various different aspects, and (c) that performance on this test is predictive for one’s ability for such tasks. Following this explanation, participants were provided with the actual test instruction and the test was administered. In contrast to the explanation group, participants in the control group were only presented with the test instruction but not the justification concerning content, job-relatedness, and predictive validity of the test. After the test, all participants had to complete a questionnaire concerning their perceptions of the test.

**Ability test.** A shortened version of d2-R test of attention (Brickenkamp, Schmidt-Atzert, & Liepmann, 2010) was used. This paper-pencil test measures selective attention and its suitability and validity for tasks that require simultaneous attention has been well-documented (cf. Brickenkamp et al., 2010). However, Kersting’s (2008) comparison of
several common German ability tests had revealed that the d2-R was perceived rather negatively.

In the d2-R, participants are presented with rows of letters and their task is to cross out any letter \(d\) that has two dashes printed below or above (i.e., \(ds\) with two dashes printed below, two dashes printed above, or one dash printed below and another one printed above are valid targets). Furthermore, each row also contains various distractor letters that are similar to the target (i.e., \(ps\) with dashes or \(ds\) with one or three dashes). Participants have 20 seconds to complete each row before they are told to carry on with the next row irrespective of how far they have proceeded in the current row.

For the current study, all participants were administered 9 rows of letters. In line with the test manual, test performance was calculated by subtracting all errors (letters that were incorrectly crossed out as well as missed targets) from the overall number of letters, which was determined by summing the number of letters completed in each row.

**Test takers’ perceptions.** We used different scales to assess test takers’ perceptions. Specifically, we measured perceived face validity (\(\alpha = .85\), e.g., “I did not understand what the test had to do with my targeted job.”) and perceived predictive validity (\(\alpha = .85\), e.g., “My performance on the test is a good indicator of my ability to do the targeted job.”) with two five-item scales that were adapted from Smither, Reilly, Millsap, Pearlman, and Stoffey (1993) and that were also used in previous research on the effects of explanations on test takers’ perceptions (e.g., Holtz et al., 2005). Overall procedural fairness (\(\alpha = .70\), e.g., “Overall, the test used was fair.”) was measured with a three-item procedural justice scale from Bauer et al. (2001). Finally, test motivation (\(\alpha = .79\), e.g., “Doing well on this test was important to me.”) was assessed with three items from Arvey, Strickland, Drauden, and Martin (1990). Five-point scales (1 = *strongly disagree* to 5 = *strongly agree*) were used for all items.

**Results**
First, we determined whether the experimental groups were comparable with regard to demographic and background variables. We found no significant differences with regard to age, gender, number of previous job applications, or previous experience with selection tests, all $t$s < 1.63, all $p$s > .10.

Table 1 shows descriptive information and correlations for all study variables. In line with previous research (Brickenkamp et al., 2010), there was a significant negative correlation between age and test performance, $r = -.26$, $p < .05$.

Means and $SD$s for the dependent variables for the two groups and effect sizes from tests of differences between the two groups are presented in Table 2. In line with our hypothesis, means for test takers’ perceptions were more positive in the explanation group. Accordingly, independent-samples $t$-tests also confirmed that the ratings were significantly higher for face validity, $t(94) = 3.72$, $p < .01$, Cohen’s $d = .76$, predictive validity, $t(94) = 3.96$, $p < .01$, Cohen’s $d = .81$, and procedural fairness $t(94) = 2.03$, $p < .05$, Cohen’s $d = .41$, for the group that had received a justification compared to the control group. However, the difference for participants’ test motivation was not significant, $t < 1$. Furthermore, there was also no significant difference between the groups with regard to their test performance, $t < 1$.

**Discussion**

In contrast to previous research (e.g., Truxillo et al., 2009), our results suggest that providing an explanation can have beneficial effects on test takers’ perceptions of an ability test that measures attention: Participants who received a justification before the test had more positive perceptions after the test with regard to face validity, predictive validity, and procedural fairness. However, the justification did not affect participants’ test motivation or test performance significantly.

Even though the present results seem promising in light of Truxillo et al.’s (2009) findings with regard to the possibility to improve applicants’ perceptions of ability tests, some caution is still necessary. Specifically, considering the specific nature of the ability test that
we used, our results cannot be taken as evidence that a justification will always suffice to improve test takers’ perceptions of other ability tests and especially of GMA tests as well. However, the question of whether our results generalize to GMA tests seems highly relevant given that there is evidence that GMA tests have better criterion-related validity (e.g., Schmidt & Hunter, 2004) than attention tests or perceptual tests in general (cf. Hunter & Hunter, 1984) and also given their higher usage. Furthermore, as already noted above, previous research also found that applicants’ perceptions of GMA tests are more favorable in comparison to personality or integrity tests (Anderson et al., 2010; Hausknecht et al., 2004). In contrast to this, the d2-R was perceived more negatively than most of the other ability tests covered in Kersting’s (2008) study. Thus, given the rather negative perceptions of the d2-R, it might be easier to improve test takers’ perceptions (at least to a certain degree) of this test than of a GMA test. Therefore, we set up Study 2 to evaluate to which degree the results from Study 1 generalize to a GMA test.

**Study 2**

**Method**

**Sample.** Altogether, 149 students from a German university took part in Study 2. However, 5 participants were excluded from the analyses because they were psychology majors beyond the first semester (even though we had informed participants from psychology beforehand that we only searched for freshmen). As in Study 1, the rationale for this was to prevent that study participants had previous training in psychometric testing. Furthermore, another participant was excluded because of insufficient knowledge of the German language. Therefore, the final sample for the Study 2 consisted of 143 participants (mean age = 20.03 years, SD = 2.50; 55% females) from various fields of study. The largest groups were majoring in computer science (27.3%), psychology (17.5%), and biochemistry (11.9%). Of
the participants, 60% had already applied for a job in the past and 23% had completed selection tests before. Additionally, 18% currently held a job.

**Procedure.** The procedures were identical to Study 1 with a few exceptions the first of which was that the justification was adjusted to the GMA test that we used. Thus, test takers in the explanation group were informed (a) that the test is an established procedure to measure cognitive abilities such as comprehension, reasoning, and problem solving, (b) that these abilities are important to quickly acquire new job knowledge and to solve new tasks and problems on the job in a swift and flexible way and (c) that performance on this test is predictive for one’s ability for such tasks. This justification was inserted before the general test instruction for the GMA test. It was printed in a frame and in bold face to increase its salience. In contrast to the explanation group, participants in the control group were only presented with the standard test instruction but not the justification concerning content, job-relatedness, and predictiveness of the test.

Again, all participants had to complete a questionnaire concerning their test perceptions once they had completed the test. Furthermore, we also included a manipulation check after the items related to test takers’ perceptions. This manipulation check asked participants to indicate whether they had received an explanation why the test they had just completed can be used for selection purposes and participants had to answer this question with *yes* or *no*. Finally, we also asked for participants’ final high school grade point average (GPA).

**Ability test.** GMA was assessed with the 10-Minute Test by Musch et al. (2011). This test is a short test that assesses GMA within 10 minutes. It is comprised of 32 items that measure both fluid (e.g., numerical sequences) and crystallized cognitive ability (e.g., “Which of the following fractions is the smallest? 1) 2/4; 2) 2/5; 3) 1/3; 4) 3/8”). Thus, similar to the Wonderlic Personnel Test (Wonderlic Inc., 2002), the items are taken from different ability domains with the aim of yielding an overall score for individuals’ GMA. In line with this, the
10-Minute Test proved to be highly g-saturated in studies by Ostapczuk, Musch, and Lieberei (2011) and Ostapczuk, Wagner, and Musch (2014). Its split-half reliability (using an odd-even split) was .76 in the present sample.

**Test takers’ perceptions.** We used the same scales as in Study 1 to assess test takers’ perceptions. The internal consistencies were .78 for perceived face validity and for perceived predictive validity, .65 for overall procedural fairness, and .73 for test motivation.

**Results**

First, we again determined whether the experimental groups were comparable with regard to demographic and background variables and also whether participants had correctly answered the manipulation check. Similar to Study 1, we did not find any significant differences with regard to age, gender, mother tongue, number of previous job applications, or high school GPA, all $t_s < 1.12$, all $p_s > .27$. However, participants in the control group ($M = 0.30$, $SD = 0.46$) had more experience with selection tests than in the explanation group ($M = 0.11$, $SD = 0.32$), $t(124.62) = 2.76, p < .05$. Furthermore, it turned out that 50 of the 143 participants answered the manipulation check incorrectly. These participants were not distributed equally among the two groups. Instead, the majority of them (i.e., 33) were from the condition that had received an explanation (initial $n = 70$) whereas only 17 were from the control condition (initial $n = 69$; cf. Table 4). Four additional participants (two from each group) did not answer the manipulation check at all and were therefore excluded from further analyses.

Table 3 shows descriptive information and correlations for all study variables and for the manipulation check. As can be seen, there were significant correlations between gender on the one hand and test motivation, $r = -.19$, test performance, $r = -.20$, and high school GPA, $r = -.33$, all $p_s < .05$. Thus, women were less motivated, showed lower test performance in the GMA test, and had a better GPA (in Germany, lower grades represent better performance). Furthermore, to our initial surprise, the correlation between GMA test scores and GPA was
not significant, $r = -.11$. However, after controlling for age and mother tongue, this correlation increased and turned significant, $r = -.22$, $p < .05$.

To test whether the explanation improved test takers’ perceptions, we conducted separate analyses for participants who answered the manipulation check correctly and those who provided an incorrect answer to it. The rational for this was that it seems unlikely that the experimental groups differ on the basis of the explanation if participants did not pay attention to this explanation. Means and $SDs$ for the dependent variables and effect sizes from statistical tests of differences between the experimental groups are presented in Table 4.

In line with our hypothesis, means for test takers’ perceptions were more positive in the explanation group when we considered data for participants with a correct answer to the manipulation check. Furthermore, independent-samples $t$-tests also confirmed that the ratings were significantly higher for face validity, $t(87) = 2.01$, $p < .05$, Cohen’s $d = .46$, and for predictive validity, $t(87) = 2.48$, $p < .05$, Cohen’s $d = .53$, for the group that had received a justification. However, for procedural fairness the means did not differ significantly, $t < 1$. As in Study 1, the difference for participants’ test motivation was not significant, $t < 1$. Finally, however, participants who had received an explanation obtained significantly lower scores in the GMA test, $t(87) = 3.69$, $p < .01$, Cohen’s $d = -.79$

In contrast to participants who answered the manipulation check correctly, test takers’ perceptions between the two groups did not differ for participants who failed the manipulation check, all $ts < 1$. To our surprise, however, the group with the explanation reported higher test motivation, $t(48) = 3.00$, $p < .01$, Cohen’s $d = .90$.

Finally, given that participants in the control group had significantly more experience with selection tests then the explanation group we also repeated the comparisons between the two groups and used experience with selection tests as a covariate. These additional analyses led to the same pattern of results. Furthermore, given the unexpected difference for performance in the GMA test among those participants who answered the manipulation check
correctly, we also repeated the analyses with mother tongue as a covariate. However, this also did not change the general pattern of results and did not reduce the obtained difference.

**Discussion**

The most important result was that explanations were also effective in improving two of the three test taker perception variables related to a GMA test – at least among those participants who answered the manipulation check correctly. However, there were two surprising results in addition to the high percentage of participants who did not read our instructions carefully. The first were the relatively low values for the test taker perception variables. These values seemed comparable to those from Study 1 even though we had not intended to choose a test that provokes negative reactions in the first place but a test that is comparable to other common tests like the Wonderlic (Wonderlic Inc., 2002). The second and even more surprising effect was that the explanation group obtained significantly lower test scores after the explanation. Thus, given that an explanation is unlikely to alter participants’ true ability level it seems as if the explanation impaired participants in performing at their true level. It might have been that the explanation increased the personal relevance of the test results and even though our study did not represent a high-stakes situation participants might have experienced higher test taking anxiety after the explanation of the test content and its relevance. Unfortunately, we did not measure anxiety in the present study to test this suggestion. However, in line with this suggestion, Hausknecht et al. (2004) found a substantial negative relationship between anxiety and performance in selection procedures.

**General Discussion**

In contrast to meta-analytic findings by Truxillo et al. (2009) and to the corresponding primary studies, the present results clearly suggest that providing an explanation can have beneficial effects on test takers’ perceptions of ability tests: Participants who received a justification before the completion of the test had more positive perceptions with regard to
face validity, predictive validity, and – at least for the attention test – also with regard to procedural fairness. However, in Study 2, these enhanced perceptions were only found for participants who were able to answer the manipulation check correctly, which confirmed that they had indeed paid attention to the explanation.

We suggest that the main reason why the present study found positive effects of an explanation in contrast to earlier research is that we used a justification that focused on the content of the test, explained the job relevance of the targeted constructs, and stressed the predictiveness of the test. Furthermore, our results suggest that this kind of explanation is not only relevant when tests seem rather abstract (as in Study 1) but also when tests with common GMA items (as in Study 2) are used. In both cases, it seems beneficial to provide test takers with an explanation of what the test is measuring and why it is related to work outcomes.

In contrast to Truxillo et al.’s (2009) meta-analytic results the explanation in our studies did not increase test motivation and in Study 2 it even impaired test performance. Concerning test motivation, it might well be that the explanation did not make much of a difference because of the relatively high motivation in the first place. Thus, the null effect for test motivation might be due to a ceiling effect. However, if the negative effect on test performance in the GMA test replicates in future research and if it is really due to increased anxiety as suggested above, then it should be explored how the relevance of the test content can be explained in an alternative way that provokes less anxiety. Alternatively, the explanation could be supplemented with an uncertainty reduction explanation. In a recent study by McCarthy, Bauer, Truxillo, Campion, et al. (2017) such an explanation was combined with an explanation about test content and the testing process as well as with a social fairness explanation that stressed respect for test takers who had to complete a work sample test afterwards. It was found that this extended explanation led to significantly improved fairness perceptions.

Limitations and Lines for Future Research
A possible objection concerning the beneficial effects on test takers’ perceptions might be related to the absolute level of the perception variables that was observed in the experimental groups: Even though the explanation improved these perceptions significantly and even though most effect sizes were in the moderate to large range, the means for the different perceptions were still close to the actual mid-point of the scale. However, this is comparable with other primary studies (e.g., Horvath et al., 2000; Rolland & Steiner, 2007).

Another limitation that concerns the generalizability of our results is our use of student samples. Previous research suggests that effect sizes in applicant reactions research differ between field studies with actual applicants on the one hand and simulated selection processes with student participants on the other hand. Future research is therefore needed to evaluate the actual effects of providing explanations to test takers’ in high-stakes selection settings. However, it seems possible that the present studies provide more conservative estimates of the actual effects of explanations in comparison to field studies with actual applicants, because additional analyses by Truxillo et al. (2009) revealed that explanations have larger effects in field studies. Furthermore, this parallels meta-analytic evidence from Hausknecht et al. (2004) who also found that applicants’ perceptions were more strongly related to many outcome measures in authentic selection contexts than in hypothetical or simulated ones. Thus, test takers in real high-stakes selection contexts for whom the test and the test outcome are more relevant than for study participants in selection simulations might be more receptive to the provision (or to the lack) of explanations.

In addition, it might also be that the high percentage of participants who did not carefully read the instructions and the explanation does not generalize to high-stakes settings where applicants are probably more anxious to not miss relevant pieces of information. However, given that it is not uncommon in laboratory research (Oppenheimer, Meyvis, & Davidenko, 2009) that substantial numbers of participants fail to read and follow the instructions, it might be that this carelessness also contributed to smaller effects of
explanations in earlier simulation studies. Again, future research is needed to evaluate this suggestion.

**Practical Implications**

Taken together, our results suggest that test administrators should explain to test takers what a given ability test is measuring and why it is important in the context of a selection process. Our hope is that providing such explanations will help to improve perceptions of ability tests in the long run. However, to avoid negative side effects on test performance, additional steps might be beneficial. Furthermore, test administrators should also try to ensure that applicants do indeed perceive the explanation.
References


### Table 1

**Descriptive Statistics and Intercorrelations of Study Variables in Study 1**

<table>
<thead>
<tr>
<th>Variables</th>
<th>M</th>
<th>(SD)</th>
<th>1.</th>
<th>2.</th>
<th>3.</th>
<th>4.</th>
<th>5.</th>
<th>6.</th>
<th>7.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Age</td>
<td>23.17</td>
<td>(3.82)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Gender(^a)</td>
<td>0.49</td>
<td>(0.50)</td>
<td>-.28**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Face validity</td>
<td>2.92</td>
<td>(0.86)</td>
<td>-.10</td>
<td>.20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Predictive validity</td>
<td>2.48</td>
<td>(0.82)</td>
<td>-.14</td>
<td>.18</td>
<td>.66**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Procedural fairness</td>
<td>3.15</td>
<td>(0.72)</td>
<td>-.01</td>
<td>.16</td>
<td>.46**</td>
<td>.51**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Test motivation</td>
<td>4.24</td>
<td>(0.56)</td>
<td>-.20</td>
<td>.21*</td>
<td>.13</td>
<td>.15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Test performance</td>
<td>143.06</td>
<td>(31.39)</td>
<td>-.26*</td>
<td>.19</td>
<td>.16</td>
<td>.07</td>
<td>.09</td>
<td>.02</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* N = 96. \(^a\) Gender was coded as 0 = male, 1 = female.  
*\(p < .05\), **\(p < .01\).
### Table 2

Means and Standard Deviations for the Two Experimental Groups in Study 1

<table>
<thead>
<tr>
<th>Variables</th>
<th>No explanation ($n = 46$)</th>
<th>Explanation ($n = 50$)</th>
<th>Effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>(SD)</td>
<td>$M$</td>
</tr>
<tr>
<td>Face validity</td>
<td>2.60</td>
<td>(0.86)</td>
<td>3.21</td>
</tr>
<tr>
<td>Predictive validity</td>
<td>2.16</td>
<td>(0.75)</td>
<td>2.78</td>
</tr>
<tr>
<td>Procedural fairness</td>
<td>2.99</td>
<td>(0.76)</td>
<td>3.29</td>
</tr>
<tr>
<td>Test motivation</td>
<td>4.22</td>
<td>(0.60)</td>
<td>4.26</td>
</tr>
<tr>
<td>Test performance</td>
<td>146.39</td>
<td>(31.17)</td>
<td>140.00</td>
</tr>
</tbody>
</table>

*Note.* Effect sizes depict Cohen’s $d$. Positive values indicate that the experimental group that had received an explanation had higher means than the no explanation control group. Asterisks represent significance levels from independent-samples $t$-tests.

* $p < .05$, ** $p < .01$. 
Table 3

Descriptive Statistics and Intercorrelations of Study Variables in Study 2

<table>
<thead>
<tr>
<th>Variables</th>
<th>M</th>
<th>(SD)</th>
<th>1.</th>
<th>2.</th>
<th>3.</th>
<th>4.</th>
<th>5.</th>
<th>6.</th>
<th>7.</th>
<th>8.</th>
<th>9.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Age</td>
<td>20.03</td>
<td>(2.50)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Gender(^a)</td>
<td>0.55</td>
<td>(0.50)</td>
<td>-0.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Face validity</td>
<td>2.98</td>
<td>(0.75)</td>
<td>-0.05</td>
<td>-0.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Predictive validity</td>
<td>2.37</td>
<td>(0.74)</td>
<td>-0.22*</td>
<td>-0.09</td>
<td>0.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Procedural fairness</td>
<td>2.99</td>
<td>(0.75)</td>
<td>-0.10</td>
<td>0.05</td>
<td>0.23</td>
<td>0.35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Test motivation</td>
<td>3.97</td>
<td>(0.74)</td>
<td>-0.05</td>
<td>-0.19</td>
<td>-0.01</td>
<td>0.15</td>
<td>0.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Test performance</td>
<td>19.79</td>
<td>(4.26)</td>
<td>-0.12</td>
<td>-0.20</td>
<td>-0.02</td>
<td>0.05</td>
<td>0.08</td>
<td>0.42</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. High school GPA</td>
<td>2.00</td>
<td>(0.63)</td>
<td>0.15</td>
<td>-0.33</td>
<td>-0.03</td>
<td>-0.10</td>
<td>-0.07</td>
<td>0.09</td>
<td>-0.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Manipulation check(^b)</td>
<td>0.64</td>
<td>(0.48)</td>
<td>0.02</td>
<td>-0.07</td>
<td>0.03</td>
<td>0.11</td>
<td>0.03</td>
<td>-0.01</td>
<td>-0.04</td>
<td>-0.02</td>
<td></td>
</tr>
</tbody>
</table>

Note. \(N = 143\). GPA = Grade point average. \(^a\) Gender was coded as 0 = male, 1 = female. \(^b\) Manipulation check was coded as 0 = incorrect, 1 = correct. * \(p < .05\), ** \(p < .01\).
### Table 4

Means and Standard Deviations for the Two Experimental Groups in Study 2

<table>
<thead>
<tr>
<th>Variables</th>
<th>Incorrect manipulation check</th>
<th>Correct manipulation check</th>
<th>Effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No explanation ($n = 17$)</td>
<td>Explanation ($n = 33$)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$M$ (SD)</td>
<td>$M$ (SD)</td>
<td></td>
</tr>
<tr>
<td>Face validity</td>
<td>2.96 (0.66)</td>
<td>2.95 (0.74)</td>
<td>-.01</td>
</tr>
<tr>
<td>Predictive validity</td>
<td>2.27 (0.84)</td>
<td>2.25 (0.79)</td>
<td>-.02</td>
</tr>
<tr>
<td>Procedural fairness</td>
<td>2.96 (0.83)</td>
<td>2.97 (0.73)</td>
<td>.01</td>
</tr>
<tr>
<td>Test motivation</td>
<td>3.55 (0.74)</td>
<td>4.19 (0.70)</td>
<td>.90**</td>
</tr>
<tr>
<td>Test performance</td>
<td>19.00 (3.94)</td>
<td>20.61 (4.18)</td>
<td>.39</td>
</tr>
</tbody>
</table>

|                    | No explanation ($n = 52$)    | Explanation ($n = 37$)      |             |
|                    | $M$ (SD)                     | $M$ (SD)                    |             |
| Face validity      | 2.86 (0.67)                  | 3.21 (0.89)                 | .46*        |
| Predictive validity| 2.27 (0.66)                  | 2.64 (0.74)                 | .53*        |
| Procedural fairness| 2.96 (0.74)                  | 3.08 (0.71)                 | .17         |
| Test motivation    | 4.01 (0.72)                  | 3.88 (0.74)                 | -.18        |
| Test performance   | 21.08 (4.06)                 | 17.81 (4.20)                | -.79**      |

*Note.* Effect sizes depict Cohen’s $d$. Positive values indicate that the experimental group that had received an explanation had higher means than the no explanation control group. Asterisks represent significance levels from independent-samples $t$-tests.  
* $p < .05$, ** $p < .01$. 