Chapter 6
Applying organizational justice theory to admission into higher education: Admission from a student perspective

This chapter was published as:
Abstract

Applicant perceptions of methods used in admission procedures to higher education were investigated using organizational justice theory. Applicants to a psychology study program completed a questionnaire about several admission methods. General favorability, ratings on justice dimensions, relationships between general favorability and these dimensions, and differences in perceptions depending on gender and on the aim of the admission procedure (selection or matching) were studied. In addition, the relationship between favorability and test performance and the relationship between favorability and behavioral outcomes were investigated. Applicants rated interviews and curriculum-sampling tests most favorably. Contrary to expectations based on the existing literature, high school grades were perceived least favorably and there was no relationship between applicant perceptions and enrollment decisions. In line with previous research in the employment literature, general favorability was most strongly related to face validity, study-relatedness, applicant differentiation, the chance to perform, perceived scientific evidence, and perceived wide-spread use. We found no differences in applicant perceptions based on gender and small differences based on the aim of admission procedures. These results extend the applicant perceptions literature to educational admission and the results are useful for administrators when choosing methods to admit students.
Applying organizational justice theory to admission into higher education

6.1 Introduction
In recent years there has been an increasing interest in the use of nontraditional instruments for admission into higher education, such as the use of personality questionnaires, motivation questionnaires, biodata, and curriculum-sampling tests (Niessen, Meijer, & Tendeiro, 2016; Schmitt, 2012; Visser, van der Maas, Engels-Freeke, & Vorst, 2012). Through the administration of these instruments as alternatives for or in addition to traditional entrance exams and high school performance, a broader set of characteristics and skills can be evaluated (e.g., Lievens & Coetsier, 2002; Schmitt, 2012; Shultz & Zedeck, 2012). Most studies have focused on the effectiveness of these instruments from the perspective of the educational institutions by studying predictive validity and differences between relevant groups. Although such studies are important and show practically and theoretically relevant results, very little attention has been paid to applicant perceptions of different admission methods. Applicant perceptions of selection methods have been mainly studied in the context of personnel selection. However, with the increasing interest in the use of different admission methods in higher education, I/O psychologists, selection officers, and other professionals are confronted with the question which methods are preferred by applicants in educational admission (e.g., Schmitt, 2012). In the present study, we tried to answer this question by investigating applicant perceptions of different admission methods in higher education, and by investigating relationships of applicant perceptions with test performance and future behavior. In addition, we studied differences in applicant perceptions and admission method preferences for male and female applicants, and differences in admission method preferences depending on the aim of the admission procedure; selection (high-stakes), or matching (low-stakes).

6.1.1 Applicant Perceptions
Applicant perceptions are “attitudes, affect, or cognitions an individual might have about a selection process” (Ryan & Ployhart, 2000, p. 566), and these perceptions have been widely studied in the context of personnel selection. Different models (Chan, Schmitt, DeShon, Clause, & Delbridge, 1997; Gilliland, 1993; Ryan & Ployhart, 2000) and instruments (Bauer, Truxillo, Sanchez, Craig, Ferrara, & Campion, 2001; Sanchez, Truxillo, & Bauer, 2000; Steiner & Gilliland, 1996) have been developed and consequences of applicant perceptions have been studied. Results showed that applicant perceptions of selection methods are related to test validity, organizational attractiveness, application recommendations to others, job-offer acceptance, litigation likelihood, applicant withdrawal, and purchase intentions (Bauer et al., 2001; Gilliland, 1994; Hausknecht, Day, & Thomas, 2004;
Many of these outcomes are, mutatis mutandis, also important for educational institutes. Moreover, higher educational institutes serve important societal purposes and the opportunity to participate in higher education has a large impact on the careers and thus future lives of individuals. Because of the impact of higher education on society and individuals, the perceptions of stakeholders to selection methods are of great importance. An example is the ongoing public debate about the content and importance of the SAT in college admissions and the recent changes made to increase relevance and face validity (e.g., Balf, 2014). Furthermore, it is not self-evident that results based on studies conducted in personnel selection contexts can be generalized to the context of admission to higher education. The outcomes to be predicted in both contexts differ; in personnel selection the main outcome to be predicted is job performance, whereas in educational selection it is academic performance. These different outcomes are predicted by partly different instruments or methods. Some instruments are used in both contexts (e.g., cognitive ability tests, personality questionnaires), but other frequently used admission methods are unique to the context of higher education (high school GPA, lottery). Furthermore, the popularity of different methods may differ across the two contexts (e.g., Ryan, McFarland, Baron, & Page, 1999).

6.1.2 Theoretical Framework

The dominant perspective on applicant perceptions of selection methods is based on organizational justice theory (Gilliland, 1993). Within organizational justice theory, several procedural justice dimensions are proposed that explain the process favorability of selection methods. Procedural justice concerns the procedures used to determine the best applicants, opposed to distributive justice, which is focused on the outcomes of the selection procedures (Steiner & Gilliland, 2001). Process favorability (a general preference for a selection method) is determined by perceived fairness and perceived predictive validity (Smither et al., 1993; Steiner & Gilliland, 1996). The seven proposed dimensions of procedural justice are scientific evidence, the right to use information, applicant differentiation, interpersonal warmth, face validity, wide-spread use, and invasion of privacy. These dimensions are usually measured with single items (Steiner & Gilliland, 1996). The organizational justice perspective was supported by findings in many studies (e.g., Schmitt, Oswald, Kim, Gillespie, & Ramsay, 2004; Smither et al., 1993). In the remainder of this chapter we shorten the terms process favorability and
procedural justice dimensions to general favorability and justice dimensions for simplicity.

Sanchez et al. (2000) proposed an alternative perspective on applicant perceptions based on expectancy theory. The three major components of expectancy theory are valence (the desirability of the outcome), instrumentality (the belief that good performance will lead to the desired outcome), and expectancy (the subjective belief that effort will increase the chance of the desired outcome). Sanchez et al. (2000) proposed that these components might partly explain test-taking motivation and procedural justice perceptions.

Another possible determinant of applicant perceptions is the self-serving bias (Chan et al., 1997; Chan, Schmitt, Sacco, & DeShon, 1998; Schmitt et al., 2004). According to this theory, applicants who perform poorly attribute those results to a lack of relevance and fairness of the test. In the studies cited above, small to moderate positive relationships were found between test scores and post-test applicant perceptions, even when controlling for pretest applicant perceptions (Chan et al., 1998).

A more specific characteristic of some methods that has received much attention but has rarely been studied in relation to applicant perceptions is the fakability or cheatability of selection methods. Many nontraditional methods that are currently receiving attention measure typical behavior (e.g., personality questionnaires, situational judgment tests (SJTs), biodata). These types of tests are susceptible to cheating or faking when used in maximum performance contexts such as selection situations (Birkeland, Manson, Kisamore, Brannick, & Smith, 2006; Viswesvaran & Ones, 1999). Some studies showed that the perceived fakability of methods was also related to applicant perceptions (Gilliland, 1995; Schreurs, Derous, Proost, Notelaers, & De Witte, 2008).

### 6.1.3 Applicant Perceptions in Personnel Selection

Many studies on applicant perceptions have been conducted in the context of personnel selection (e.g., Anderson & Witvliet, 2008; Gilliland, 1994; Smither et al., 1993). Anderson, Salgado, and Hülsheger (2010) conducted a meta-analysis on applicant perceptions using data from many different countries. They found that applicant perceptions were generalizable across specific selection situations and countries. In general, work samples and interviews were the most favorable methods, résumés, cognitive ability tests, references, biodata, and personality questionnaires were rated favorably, and honesty tests, personal contacts, and graphology were rated least favorably. Anderson et al. (2010) also found that for
the more specific justice dimensions, work samples and interviews were perceived as highly face-valid and were rated favorably on most dimensions. However, work samples were rated slightly lower on interpersonal warmth, scientific evidence, and widespread use. Cognitive ability tests were rated highest for invasion of privacy, and personality tests and biodata were rated moderately on most dimensions.

Relationships between ratings on the justice dimensions and general favorability have been studied to gain insight in the determinants of applicant perceptions. The results were mostly consistent across studies and showed that face validity, applicant differentiation and widespread use were strongly related to general favorability, the right to use and scientific evidence were moderately related to general favorability, and interpersonal warmth and invasion of privacy showed small relations to general favorability (Bertolino & Steiner, 2007; Ispas, Ilie, Iliescu, Johnson, & Harris, 2010; Moscoso & Salgado, 2004; Nikolaou & Judge, 2007; Steiner & Gilliland, 1996). Another dimension that was strongly related to general favorability but that was not included in Steiner and Gilliland’s (1996) framework was job-relatedness (Bauer et al., 2001). In conclusion, high-fidelity methods (methods that are similar to the criterion in content) like work samples, and methods that make applicants feel that they can show their unique skills and abilities, like interviews, are perceived favorably by applicants (e.g., Ployhart, Schneider & Schmitt, 2006).

6.1.4 Applicant Perceptions in Higher Education
In the context of higher education, few studies on applicant perceptions of admission methods have been conducted, and the available studies only evaluated specific admission instruments and specific aspects of applicant perceptions. Patterson, Zibarras, Carr, Irish, and Gregory (2011) found that applicants to a postgraduate medical training program rated a clinical problem-solving task as significantly more relevant than a SJT, and a simulated patient task as significantly more relevant than a group exercise and a written exercise. Lievens (2013) found that medical school applicants rated an SJT measuring interpersonal skills as significantly more face valid than cognitive science knowledge tests. These results showed that methods that matched the context of the programs were rated more positively than more general or lower-fidelity methods (Kluger & Rothstein, 1993; Ployhart et al., 2006).

In contrast, Schmitt et al. (2004) studied fairness and relevance perceptions of undergraduate students to SAT/ACT scores, and a combined biodata/SJT instrument designed to predict broad college student performance criteria. They found that fairness perceptions for SAT/ACT were higher than for the SJT and
biodata instruments, and that fairness ratings were low for the latter two methods. There were no significant differences between the methods for perceived relevance. Schmitt et al. (2004) also studied the effect of direct or indirect self-serving bias and found that perceived performance was positively related to perceptions of relevance, which in turn were positively related to fairness perceptions. Finally, Schmitt (2012) discussed that their “previous collection of reactions measures suggests that students view HSGPA as the most appropriate index of student potential with the use of biodata, SJT, and SAT/ACT less favorably viewed. The latter three indices were perceived to be about equally relevant and fair” (p. 28).

6.1.5 Potential Variables Affecting Applicant Perceptions
It is well known that in admission to higher education, performance on some commonly used predictors differs across males and females, and that some predictors show differential prediction by gender (Fischer, Schult, & Hell, 2013; Keiser, Sackett, Kuncel, & Brothen, 2016). Males tend to obtain higher scores on cognitive tests than females, and female academic performance tends to be slightly underpredicted by scores on cognitive tests such as the SAT and ACT (Fischer et al., 2013). Conversely, females tend to score higher on relevant personality constructs such as conscientiousness, procrastination (reversed), and academic skills (Keiser et al., 2016). Therefore, applicant perceptions of admission methods may differ for males and females.

Furthermore, the selection ratio of universities can differ widely. Some admission procedures are aimed at strict selection and thus admission of the best applicants, while other procedures are aimed at determining student-program fit (matching), resulting in an enrollment advice. Applicant perceptions of admission methods may differ depending on the aim of admission procedures. Some methods may be perceived more favorably when they are used to determine which applicants would be the most successful students (selection), while others may be perceived more favorably when they are used to gain insight in applicants’ fit to a program (matching).

6.1.6 Aims of the Present Study
Educational institutes can often choose their own admission methods and criteria to select students, and there is wide variety of possible methods and instruments. Knowledge about perceptions of applicants to higher education about these methods is lacking, and through this study we aimed to fill this gap. Educational institutes can then take this information into account in designing their admission policies. In addition, we investigated if results based on organizational justice
theory obtained in an educational context and applied to educational admission methods are comparable to results obtained in personnel selection contexts. We also investigated if applicant perceptions differed depending on gender or on the aim of admission procedures.

After a long tradition of open admission and lottery admission, selective admission was recently implemented in the Netherlands. We studied applicant perceptions of methods that are often used or suggested in the literature, or have recently been implemented in admission to Dutch higher education, based on inspection of websites of higher education institutions (ISO, 2014). These methods were cognitive ability tests, personality questionnaires, motivation questionnaires, biodata, high school grades, skills tests, curriculum-sampling tests, interviews, and lotteries. Table A6.1 in the Appendix provides a brief description of each method.

First, we studied the general favorability of the admission methods in a selection and a matching sample. We hypothesized that interviews and high-fidelity methods like curriculum-sampling tests and skills tests would be perceived as most favorable, followed by cognitive ability tests, high school grades, and biodata; lotteries would be perceived as least favorably. Second, we studied ratings on several justice dimensions for each of the methods and their relationships with general favorability to gain insight in determinants of applicant perceptions in higher education. Third, we examined whether applicants perceptions of admission methods differed based on gender, and on the aim of the admission procedure (selection or matching). Fourth, we studied the relationships between general favorability of skills tests and curriculum-sampling tests, and actual test scores obtained with these methods. On the basis of the self-serving bias theory we expected that applicants with lower scores would rate the methods as less favorably. Finally, we tried to replicate the relationship between applicant perceptions and behavioral outcomes such as job-offer acceptance found in personnel selection contexts (Hausknecht et al., 2004; Macan et al., 1994). We analyzed the relationship between applicant perceptions of the methods used in an admission procedure and enrollment decisions and we asked the applicants if they took the admission method into account when choosing a university and a study program.

6.2 Method

6.2.1 Participants

Selection sample

The sample consisted of 220 applicants who applied to an undergraduate psychology program at a Dutch university in 2015. Before participating in the
study, the applicants participated in a selection procedure consisting of two curriculum-sampling tests and a skills test in mathematics. The curriculum-sampling tests mimicked future study behavior. For the first curriculum-sampling test applicants were asked to study two chapters of introductory psychology material, and for the second curriculum-sampling test applicants were instructed to view a video lecture. Both tests consisted of multiple-choice questions about the material. The math test consisted of items about high-school algebra and skills related to basic statistics. The selection committee rejected none of the applicants. However, the students did not know this in advance and perceived the selection tests as high-stakes tests. In addition, 134 of the 220 participants (61%) also voluntarily completed personality and motivation questionnaires for research purposes before participating in the selection procedure. After the selection procedure all applicants were asked to complete an online questionnaire about different selection methods. Participation was voluntary, 34% of all applicants completed the questionnaire. Some participants completed the questionnaire after receiving their scores (22% of the participants). Participants applied to a Dutch-spoken program (34% of the participants, 35% in the applicant pool) or to an English-spoken program. For this latter program mostly international students applied, of which 98% had a European nationality. In the group of participants, 75% was female (70% in the applicant pool). The mean age for the participants was \( M = 20 \) (\( SD = 2.3 \)), and in the total applicant pool the mean age was \( M = 20 \) (\( SD = 2.2 \)). Ten percent of the participants decided not to enroll in the program after acceptance to the program (27% in the applicant pool).

**Matching sample**

The sample consisted of 133 applicants who applied to the same undergraduate psychology program at a Dutch university in 2016. The faculty had discontinued selective admission and implemented a matching procedure instead, that consisted of the same curriculum-sampling tests as in 2015. In addition, the math test was replaced by another curriculum-sampling test about statistics, which covers a significant proportion of the curriculum. The matching procedure was aimed at helping the applicants gain insight into their fit to the program. The applicants knew that they could not be rejected, but that they would be advised about their enrollment based on the scores on the admission tests. After completing the matching tests all applicants were asked to complete an online questionnaire about different admission methods. Participation was voluntary, 29% of all applicants completed the questionnaire. Participants applied to a Dutch-spoken program (51% of the participants, 47% in the applicant pool) or to an English-spoken program. For this latter program mostly international students applied, of
which 86% was from Europe. In the group of participants, 71% was female (70% in the applicant pool). The mean age for the participants was $M = 20$ ($SD = 3.7$), and in the total applicant pool the mean age was $M = 20$ ($SD = 2.9$).

### 6.2.2 Measures
Participants completed an online questionnaire about all admission methods listed in Table A6.1 in the Appendix. For the matching sample, lottery admission was not included because it would not apply. The order of presenting the methods to the respondents was randomly generated for each respondent. Each method was briefly described, sometimes including an example item. Next, 13 items were administered that were mostly based on the questionnaire by Steiner and Gilliland (1996). The first two items (perceived predictive validity and perceived fairness) measured general favorability, and lead to an overall description of the favorability of the methods. In addition, the seven items from this questionnaire measuring justice dimensions were included. We extended the questionnaire with an item about study-relatedness, and an item about the chance to perform based on Bauer et al. (2001), a question about effort expectancy from Sanchez et al. (2000), and a question about the ease of cheating. The complete questionnaire can be found in Table A6.2 in the Appendix. Each response was provided on a seven-point scale (scored 1–7) with verbal anchors. The respondents completed the questionnaire in Dutch when they applied for the Dutch-spoken program and in English when they applied for the English-spoken program. In addition, we also asked the participants whether the selection or matching procedure used by a particular university influenced their application for a university and study program (yes, somewhat, or no). Test performance, enrollment in the program, and gender were obtained through the university administration. Informed consent was obtained from all participants to access their test scores and academic records and to match these scores and records with their responses on the questionnaire.

### 6.2.3 Procedure
For both samples, general favorability of each method for each respondent was calculated as the mean score on the two general favorability items. These mean scores were used to calculate the mean favorability and a 95% confidence interval for each selection method. The items measuring interpersonal warmth were reverse scored to ease interpretation. Mean scores and confidence intervals for the justice dimension items were also computed for all admission methods. To study relationships between general favorability and the justice dimensions, we calculated the correlation between scores on the dimension items and the mean general favorability score for each method. To investigate self-serving bias, we computed correlations between the admission test scores and the general
favorability ratings of the corresponding method. A logistic regression analysis was conducted with enrollment to the program as the dependent variable and the favorability ratings of curriculum-sampling tests and skills tests as the independent variables, based on the data obtained in the selection sample. There were 0.4% missing values in the data of the selection sample and no missing data in the matching sample. Since the percentage of missing values was very small and no patterns emerged in the missing data, we made the assumption that the data were missing completely at random and we used pairwise deletion for all analyses. To study if applicant perceptions differed depending on the aim of the admission procedure and the gender of the applicants, a repeated measures ANOVA was conducted on a dataset containing data from both samples, with the mean general favorability rating as the dependent variable, method as a within-subjects independent variable, and aim and gender as between-subjects independent variables, including an interaction term between method and aim and method and gender.

6.3 Results

6.3.1 General Favorability
First, we assessed if there were differences between participants who completed the questionnaire before or after receiving their admission scores, and between participants who did and who did not complete the personality and motivation questionnaires in the selection sample. We found no differences in favorability ratings between participants who completed the questionnaire before or after receiving their scores on the methods used in the admission procedure, with Cohen’s $d = .01, t(218) = 0.08, p = .93$ for curriculum-sampling tests, and Cohen’s $d = .20, t(218) = 1.24, p = .22$ for skills tests. We also found no differences in favorability ratings of personality questionnaires and motivation questionnaires between respondents who completed these instruments and respondents who did not, with Cohen’s $d = .07, t(217) = .47, p = .64$ for personality questionnaires, and Cohen’s $d = .14, t(217) = .96, p = .34$ for motivation questionnaires. Given these results, we combined all cases as a single selection sample.

Table 6.1 shows descriptive statistics of the general favorability ratings of each method in both samples. In the selection sample, interviews and curriculum-sampling tests received the highest ratings, with non-overlapping confidence intervals with other methods. Cognitive ability tests, skills tests, biodata, motivation questionnaires, and personality questionnaires were rated less favorably, but the confidence intervals were above or included the neutral mid-point of the scale. High school grades and lotteries were rated least favorably, with non-overlapping confidence intervals with ratings of the other methods or with the
midpoint of the scale. The results in the matching sample were similar but showed a slightly different ordering, with interviews, skills tests, curriculum-sampling tests, and cognitive ability tests rated most favorably, followed by motivation questionnaires, personality questionnaires, and biodata. High school grades were rated least favorably, with non-overlapping confidence intervals with other methods. The most salient result in both samples was the low rating of the use of high school grades. Although frequently used and strongly related to academic performance, students did not perceive high school grades as a favorable basis for admission decisions.

Table 6.1
Mean scores, standard deviations, and 95% confidence intervals for general favorability ratings obtained in the selection and the matching sample, and Cohen’s d for the difference in ratings between the matching sample and the selection sample.

<table>
<thead>
<tr>
<th>Method</th>
<th>Selection</th>
<th>Matching</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>95% CI</td>
</tr>
<tr>
<td>Interviews</td>
<td>5.29</td>
<td>1.12</td>
<td>5.15, 5.45a</td>
</tr>
<tr>
<td>Curriculum-sampling tests</td>
<td>5.16</td>
<td>1.05</td>
<td>5.03, 5.31a</td>
</tr>
<tr>
<td>Cognitive ability tests</td>
<td>4.72</td>
<td>1.21</td>
<td>4.56, 4.89b</td>
</tr>
<tr>
<td>Skills tests</td>
<td>4.70</td>
<td>1.16</td>
<td>4.53, 4.84b</td>
</tr>
<tr>
<td>Biodata</td>
<td>4.39</td>
<td>1.40</td>
<td>4.22, 4.59b,c</td>
</tr>
<tr>
<td>Motivation questionnaires</td>
<td>4.15</td>
<td>1.50</td>
<td>3.96, 4.36c,d</td>
</tr>
<tr>
<td>Personality questionnaires</td>
<td>3.81</td>
<td>1.40</td>
<td>3.64, 4.01d</td>
</tr>
<tr>
<td>High school GPA</td>
<td>3.28</td>
<td>1.38</td>
<td>3.11, 3.47e</td>
</tr>
<tr>
<td>Lottery</td>
<td>3.06</td>
<td>1.29</td>
<td>2.89, 3.23e</td>
</tr>
</tbody>
</table>

Note. Letters in superscript show overlapping confidence intervals. * p < .05

6.3.2 Justice Dimensions
Figure 6.1 shows the scores on all dimensions for each method based on the selection sample. The dimensions right to use and wide-spread use showed very small differences between the methods. For invasion of privacy there were also few differences, and none of the methods were rated as invasive. The dimensions that showed most variation between methods were interpersonal warmth, applicant differentiation, ease of cheating, effort expectancy, and chance to perform. None of
the methods were rated highly on *study-relatedness*, with the highest mean ratings around the midpoint of the scale. Curriculum-sampling tests scored high on most positive dimensions, but were also perceived as impersonal. Interviews also scored high on most positive dimension and were perceived as personal, as expected. Curriculum-sampling tests and cognitive ability tests scored highest on *scientific evidence*. Lotteries received the lowest scores on all positive dimensions, but also scored low on *ease of cheating*. The most salient results were, again, the unexpected low ratings for high school grades and the mid-range scores for curriculum-sampling tests on *chance to perform*, *study-relatedness* and *applicant differentiation*, which were lower than expected. Nontraditional measures often used to measure noncognitive skills were rated highest on *ease of cheating*, but were rated favorably for *interpersonal warmth*.

Figure 6.1. Mean scores and 95% confidence intervals for each method on each dimension in the selection sample, ordered from high to low ratings. 
*Note.* SE = Scientific evidence, FA = Face validity, AD = Applicant differentiation, IW = Interpersonal warmth, RU = Right to use, IP = Invasion of privacy, WU = Wide-spread use, SR = Study-relatedness, CP = Chance to perform, EX = Effort expectancy, EC = Ease of cheating.
Figure 6.2 displays the correlations between the dimension ratings and general favorability for all methods. The ratings on face validity were most strongly related to general favorability and this relationship was large for all methods. Other strong relationships with general favorability were found for study-relatedness, applicant differentiation, chance to perform, scientific evidence, and wide-spread use. Right to use, interpersonal warmth, and effort expectancy showed small positive or no relationships with general favorability, and these relationships varied across methods. As expected, invasion of privacy showed negative relationships with general favorability, but these relationships were mostly small and not statistically significant. A notable result was the negative correlation between effort expectancy and general favorability for personality questionnaires and motivation questionnaires. This may be explained by the possibility of faking on these methods. The dimension ease of cheating showed varying relationships with general favorability across methods. Especially motivation questionnaires were rated less favorably when they were rated as easier to fake, as were personality tests and interviews. Previous findings that face validity and job/study-relatedness were strongly related to general favorability were thus replicated. The same analyses were also conducted for the matching sample and showed very similar results (not shown). The most notable differences were seen in the ratings on study-relatedness. Skills tests were rated as most study-related in the matching sample, while they were ranked sixth on study-relatedness in the selection sample. Cognitive ability tests were rated as most study-related in the selection sample, while they were ranked seventh on study-relatedness in the matching sample. Detailed results for the matching sample can be obtained upon request.

6.3.3 Differences in Applicant Perceptions Based on Aim and Gender
Table 6.1 shows descriptive statistics for general favorability ratings of the admission methods for selection and matching purposes and Table 6.2 shows descriptive statistics for males and females in both samples. A repeated measures ANOVA was conducted to investigate if there were differences in favorability ratings depending on the aim of the admission procedure and depending on the gender of the applicants. Mauchly’s test showed that sphericity was violated with $\varepsilon > .75$, so the Huynh-Feldt correction was applied (Field, 2013). There was a small interaction effect between method and aim ($F(6.19, 2135.70) = 4.92, p < .01, \eta^2_p = .01$) and a small main effect for aim ($F(1, 345) = 7.62, p = .01, \eta^2_p = .02$), with lower favorability ratings when the aim was matching, compared to selection. The main effect for method was large ($F(6.19, 2135.70) = 81.38, p < .01, \eta^2_p = .19$).

There was also a small interaction effect between method and gender ($F(6.19, 2135.70) = 2.74, p = .01, \eta^2_p = .01$), but no main effect for gender ($F(1, 345) = 1.87, p = .18$,
When inspecting Cohen's $d$s shown in Table 6.1, we can observe that there were almost no differences in favorability ratings between the two aims, except for curriculum-sampling tests, biodata, and interviews, which were all rated less favorably when the aim was matching, with small to moderate effect sizes. Cohen's $d$s displayed in Table 6.2 showed similar results. The only method that showed a small difference in favorability based on gender was the motivation questionnaire, receiving higher favorability ratings by females than by males.

Figure 6.2. Correlations between dimension ratings and general favorability for each method based on the selection sample. Note. The dimensions are ordered according to the magnitude of the mean correlation of general favorability. If $r > .13$, $p < .05$.

6.3.4 Applicant Perceptions and Test Scores
In the selection sample we found a positive correlation between the favorability scores of the curriculum-sampling tests and the scores on the literature-based curriculum-sampling test $r = .15$ ($p = .02$). For the video-lecture curriculum-sampling test we found $r = .22$ ($p < .01$), and the correlation between favorability of
skills tests and the score on the math test was $r = .19 \ (p = .01)$. In the matching samples the math test was replaced by a curriculum-sampling test in statistics for the social sciences. The correlations between the general favorability rating of curriculum-sampling tests and test scores were $r = .26 \ (p < .01)$ for the literature-based curriculum-sampling test $r = .38 \ (p < .01)$ for the video-based curriculum-sampling test, and $r = .12 \ (p = .17)$ for the statistics curriculum-sampling test. So, in general, test scores were positively related to general favorability for that same method, but the effect sizes were small.

Table 6.2

<table>
<thead>
<tr>
<th>Method</th>
<th>Males Aim</th>
<th>Males SD</th>
<th>Males 95% CI</th>
<th>Females Aim</th>
<th>Females SD</th>
<th>Females 95% CI</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interviews</td>
<td>S</td>
<td>5.30</td>
<td>1.21</td>
<td>4.97, 5.63</td>
<td>5.30</td>
<td>1.10, 5.47</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>4.96</td>
<td>1.19</td>
<td>4.58, 5.35</td>
<td>4.89</td>
<td>1.12, 5.12</td>
<td>-.06</td>
</tr>
<tr>
<td>Curriculum-sampling tests</td>
<td>S</td>
<td>5.08</td>
<td>1.23</td>
<td>4.75, 5.41</td>
<td>5.19</td>
<td>0.98, 5.35</td>
<td>.11</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>4.40</td>
<td>1.42</td>
<td>3.94, 4.86</td>
<td>4.76</td>
<td>.96, 5.16</td>
<td>.32</td>
</tr>
<tr>
<td>Cognitive ability tests</td>
<td>S</td>
<td>4.82</td>
<td>1.26</td>
<td>4.48, 5.17</td>
<td>4.69</td>
<td>1.18, 5.11</td>
<td>-.11</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>4.74</td>
<td>1.14</td>
<td>4.37, 5.11</td>
<td>4.55</td>
<td>1.13, 4.73</td>
<td>-.17</td>
</tr>
<tr>
<td>Skills tests</td>
<td>S</td>
<td>4.75</td>
<td>1.20</td>
<td>4.42, 5.05</td>
<td>4.67</td>
<td>1.17, 4.94</td>
<td>-.07</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>4.63</td>
<td>1.31</td>
<td>4.20, 5.05</td>
<td>4.82</td>
<td>1.05, 5.04</td>
<td>.17</td>
</tr>
<tr>
<td>Biodata</td>
<td>S</td>
<td>4.16</td>
<td>1.31</td>
<td>3.81, 4.52</td>
<td>4.49</td>
<td>1.42, 4.71</td>
<td>.24</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>3.91</td>
<td>1.40</td>
<td>3.46, 4.36</td>
<td>3.74</td>
<td>1.20, 3.99</td>
<td>-.14</td>
</tr>
<tr>
<td>Motivation questionnaires</td>
<td>S</td>
<td>3.77</td>
<td>1.60</td>
<td>3.34, 4.21</td>
<td>4.29</td>
<td>1.44, 4.51</td>
<td>.35*</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>3.72</td>
<td>1.35</td>
<td>3.28, 4.16</td>
<td>4.32</td>
<td>1.28, 4.59</td>
<td>.46*</td>
</tr>
<tr>
<td>Personality questionnaires</td>
<td>S</td>
<td>3.70</td>
<td>1.64</td>
<td>3.26, 4.15</td>
<td>3.86</td>
<td>1.32, 4.07</td>
<td>.11</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>3.95</td>
<td>1.32</td>
<td>3.52, 4.38</td>
<td>3.98</td>
<td>1.24, 4.23</td>
<td>.02</td>
</tr>
<tr>
<td>High school GPA</td>
<td>S</td>
<td>3.38</td>
<td>1.55</td>
<td>2.96, 3.80</td>
<td>3.26</td>
<td>1.32, 3.64</td>
<td>-.09</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>3.14</td>
<td>1.56</td>
<td>2.63, 3.65</td>
<td>3.07</td>
<td>1.24, 3.33</td>
<td>-.05</td>
</tr>
<tr>
<td>Lottery</td>
<td>S</td>
<td>2.93</td>
<td>1.31</td>
<td>2.57, 3.28</td>
<td>3.10</td>
<td>1.28, 3.30</td>
<td>.13</td>
</tr>
</tbody>
</table>

*Note. Methods are displayed in the order of general favorability ratings. S = Selection aim, M = matching aim. \* $p < .05$
6.3.5 Applicant Perceptions and Behavioral Outcomes

Participants were asked if the selection method influenced their choice of a university and study program. For choosing a university, 20% responded that the selection method influenced their choice, 20% responded that it influenced their choice somewhat, and 60% indicated that it was of no influence. With respect to study program choice, 12% answered that the selection method was of influence, 18% answered that it influenced the choice somewhat, and 70% that it was of no influence. In the matching sample, 8% of the respondents indicated that the matching procedure influenced their choice of a university, 14% reported some influence, and 78% said that it was of no influence for choosing a university. For choosing a program, 5% indicated that the matching procedure influenced their choice, 24% report some influence and 71% report no influence.

Based on the data obtained in the selection sample, a logistic regression analysis was conducted to predict enrollment in the program based on the general favorability ratings of curriculum-sampling tests and skills tests, since these tests were used in the admission procedure. For curriculum-sampling favorability, the mean rating of applicants who did not enroll was $M = 5.4$ ($SD = 0.98$), and for applicants who did enroll the mean rating was $M = 5.1$ ($SD = 1.05$). For skills test favorability the mean rating of applicants who did not enroll was $M = 4.7$ ($SD = 0.97$), and for applicants who did enroll the mean rating was $M = 4.7$ ($SD = 1.20$). The logistic regression model with the favorability ratings as independent variables did not significantly predict enrollment (model $\chi^2(2) = 1.02$, $p = .60$), with $OR = 0.78$ (95% CI [0.48; 1.28], Wald $\chi^2 = 0.97$, $p = .33$) for general favorability of curriculum-sampling tests and $OR = 1.07$ (95% CI [0.72; 1.60], Wald $\chi^2 = 0.11$, $p = .75$) for skills tests.

6.4 Discussion

The aim of this study was to investigate applicant perceptions of admission methods used in higher education. We found some surprising results; the low favorability of using high school grades for matching or selection purposes was most surprising. High school grades are widely used in many countries and are a highly valid predictor of academic performance in higher education (e.g., Richardson, Abraham, & Bond, 2012). The low favorability of high school grades was contrary to the results found in the personnel selection literature that actual predictive validity was related to general favorability (Anderson et al., 2010), and contrary to Schmitt’s (2012) report that high school GPA was viewed most favorably by students and other stakeholders. A possible explanation for our results supported by organizational justice theory (Gilliland, 1993) and expectancy theory (Sanchez et al., 2000) is that high school grades are already obtained and cannot be altered, which may evoke feelings of “not being in control” of the
admission process. High school grades were rated low on chance to perform, applicant differentiation, and face validity, which were strongly related to general favorability. The same rationale may apply to the low favorability ratings of lotteries, which were rated least favorably on general favorability and the majority of the justice dimensions.

We also found that the nontraditional methods used to measure noncognitive characteristics (personality and motivation questionnaires and biodata) were not rated very favorably, and significantly less favorably than interviews and curriculum-sampling tests. These methods were perceived as easy to cheat and the perceived ease of cheating was negatively related to the general favorability of these methods. Effort expectancy showed a negative correlation with general favorability for motivation questionnaires and personality questionnaires, while it was positively related to general favorability for all other admission methods. This negative correlation may also be related to the possibility of faking on these methods, where “investing effort” on these methods may have been interpreted as faking by the applicants.

Because of the consistent findings of differences between males and females in scores on cognitive tests and some personality trait measures, we hypothesized that applicant reactions to admission methods may also differ between male and female applicants. We found a significant interaction effect between method and gender, but the effect size was small. In addition, we expected that the aim of the admission procedure (selection or matching) could influence applicant perceptions as well. Our results showed small significant effects for aim and for the interaction between method and aim. Applicants tended to rate most methods less favorably when the aim was matching, but these effects were small. A notable finding was that the two most favorably rated methods, interviews and curriculum-sampling tests, showed relatively large differences in favorability for the selection and matching samples. An explanation for this finding could be that the results of matching procedures are not binding, but that curriculum-sampling tasks and the interviews would require preparation and effort. When applicants have to put effort into a task that does not really have consequences, the result may be a lower appreciation of such a task than when the results would have important consequences.

With respect to the relationships between applicant perceptions and behavioral outcomes there were significant but small correlations between test performance and favorability. In contrast to findings obtained in employment settings (Hausknecht et al., 2004), we found no relationship between applicant perceptions and enrollment decisions. However, these applicants went through an admission
procedure consisting of curriculum-sampling tests and a math skills test, which were rated favorably. These results might have been different when other, less favorably rated methods were used. Although the majority of applicants in both samples indicated that the admission methods did not influence their choice of a program or a university, between 20 and 40 percent of the applicants indicated that the admission methods influenced their choice at least to some extent. These numbers could be of practical significance to higher education institutions.

6.4.1 Limitations
One limitation of this study was that we used two cohorts of applicants to a psychology program at a Dutch university, and that not all applicants participated in the study. However, the participants seemed to be representative for the entire applicant pools, with enrollment rate as an exception. The percentage of participants that chose to enroll in the program was larger than the percentage in the applicant pool. Second, the respondents did not have experience with all admission methods in the questionnaire. In the selection sample, all respondents took curriculum-sampling tests and a skills test, and some respondents also completed a personality and motivation questionnaire for research purposes, but respondents may have differed in the amount of experience they had with other methods. This may have resulted in differences in perceptions between respondents. We did not, however, find differences in perceptions between respondents who did and those who did not complete the personality and motivation questionnaires. In the matching sample, applicants only took three curriculum-sampling tests.

Another possible limitation may be that we used Steiner and Gilliland’s (1996) questionnaire that consists of single-item measures for the justice dimensions. Whereas this may lead to reduced validity when measuring broad constructs, single-item measures are suitable for narrow and specific constructs, such as the justice dimensions (e.g., Gardner, Cummings, Dunham, & Pierce, 1998). Also, Jordan and Turner (2008) found that single-items functioned well in measuring organizational justice.

6.4.2 Theoretical Implications
This study showed that organizational justice theory can be applied to applicant perceptions in an educational context, and this was the first study that applied this theory to a wide variety of admission methods in an educational context. To some extent, we found results similar to the results in personnel selection (e.g., Anderson et al., 2010), with the highest ratings for interviews and curriculum-sampling tests (a method similar to work sample tests in personnel selection). The favorability of admission methods was most strongly related to their face validity,
study-relatedness, applicant differentiation, the chance to perform, perceived scientific evidence, and perceived widespread use. In line with previous findings, we found that high-fidelity methods such as curriculum-sampling tests were rated more favorably than low-fidelity methods such as personality questionnaires. An exception was cognitive ability tests, which is not a very high-fidelity method, but was rated favorably. An explanation for the high favorability of high-fidelity methods is their high face validity and criterion-relatedness (Ployhart et al., 2006). However, curriculum-sampling tests were not rated highly on study-relatedness in this study.

While organizational justice theory could provide meaningful insight into the favorability of admission methods, the justice dimensions right to use, interpersonal warmth and invasion of privacy that were part of the original applicant perceptions scale by Steiner and Gilliland (1996) showed very little variation in ratings across methods or small correlations with general favorability. These findings are in line with studies conducted in personnel selection contexts (Bertolino & Steiner, 2007; Ispas et al., 2010; Moscoso & Salgado, 2004; Nikolau & Judge, 2007; Steiner & Gilliland, 1996), although right to use was more strongly related to general favorability in those studies. In addition, the dimensions study-relatedness and chance to perform that we included in the questionnaire used in this study, but obtained from a different instrument developed to measure procedural justice (Bauer et al., 2001) showed strong relationships with general favorability. The effort expectancy dimensions obtained from Sanchez et al., 2000 did not show such a relation. Furthermore, some dimensions that were not included in the original framework may be specifically relevant for some methods. Ease of cheating was not related to general favorability for most methods, but it was for self-report instruments. These results indicate the need for reconsidering the procedural justice dimensions that determine the general favorability of admission methods used in education, and perhaps in personnel selection as well.

6.4.3 Practical Implications
Applicant perceptions may be taken into account when choosing methods to admit students. However, they should be carefully weighted with the predictive validity of the individual selection methods. The high favorability of interviews, for example, is not in accordance with the often-found low validity and reliability of interviews, especially when they are unstructured (Dana, Dawes, & Peterson, 2013; Schmidt & Hunter, 1998). Also, the low favorability of using high school grades does not correspond with their high predictive validity (e.g., Richardson et al., 2012). When methods show similar predictive validity, the method associated with more positive applicant perceptions may be preferred. For example, Niessen et al. (2016) reported high and similar predictive validities for a curriculum-
sampling test and high school grades for first-year academic performance. Considering the high favorability of curriculum-sampling tests and the negative applicant perceptions toward using high school grades as an admission criterion, using a curriculum-sampling test may be a viable alternative to high school grades. Alternatively, interventions could be implemented to explain the use of unpopular admission methods so as to influence applicant perceptions (e.g., Truxillo et al., 2009). Furthermore, more studies that examine the behavioral consequences of positive and negative applicant perceptions are needed.
New Rules, New Tools: Predicting academic achievement in college admissions

Susan Niessen