

Kroll, Esther; Ziegler, Matthias

Article — Accepted Manuscript (Postprint)

Discrimination due to Ethnicity and Gender: How susceptible are video-based job interviews?

International Journal of Selection and Assessment

Suggested Citation: Kroll, Esther; Ziegler, Matthias (2016) : Discrimination due to Ethnicity and Gender: How susceptible are video-based job interviews?, International Journal of Selection and Assessment, ISSN 1468-2389, Wiley-Blackwell, Oxford, Vol. 24, Iss. 2, pp. 161-171, <https://doi.org/10.1111/ijsa.12138>

This Version is available at:

<https://hdl.handle.net/10419/194858>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Discrimination due to ethnicity and gender: How susceptible are video-based job interviews?

Esther Kroll and Matthias Ziegler

Department of Psychology, Humboldt Universität zu Berlin, Unter den Linden 6, 10099
Berlin, Germany.

zieglema@hu-berlin.de

Abstract

Fairness towards job applicants differing in gender and ethnicity in a video-based assessment interview was explored. For this purpose, 103 female and 105 male participants, including 38 who declared to have a migration background of their own, rated a behavior anchored rating scale after having watched the videotaped answers of a potential applicant. The domains assessed were communication skills and the capacity to work in a team. The videos of the applicants were generated with the help of standardized scripts and semi-professional actors. Eight videos were made operationalizing a two (Turkish migration background – native German) by two (male – female) by two (more positive applicant answers – moderately good applicant answers) experimental design. A multivariate analysis of variance (MANOVA) revealed a small to moderate main effect only for migration background of the applicants. Subsequent ANOVAs found that in three of the four dependent variables (DV) this effect reached significance of $p < .05$. The effects were robust against consideration of the raters' agreeableness and the raters' own migration background as covariates. Applicants with Turkish background scored higher in the evaluation of their videotaped answers than German native applicants did. Social Identity Theory (Tajfel & Turner, 1986) provides an approach to integrate these findings.

Keywords: *fairness, adverse impact, video based assessment, ethnicity, HR selection process*

Discrimination due to ethnicity and gender: How susceptible are video-based job interviews?

In the United States a great area of research focuses on discrimination based on ethnicity or gender (e.g. Dean, Roth, & Bobko, 2008; Kraiger & Ford, 1985; Roth, Huffcutt, & Bobko, 2003; Sackett & DuBois, 1991). A frequently used concept in selection contexts is “adverse impact”, which is the odds ratio of the favored, least rejected group and the disadvantaged, most rejected group (Outtz & Newman, 2010; Hattrup & Roberts, 2010). An odds ratio is considered to document an adverse impact, when the selection rate of one group is less than 80% of the group with the highest selection rate. It is important to keep in mind that subgroups’ differences are well known in research of adverse impact (Ployhart & Holtz, 2008). Therefore, evidence found in one study cannot be transferred to another subgroup or even to another country. While in the United States usually white Caucasians are contrasted to African Americans or Hispanics (e.g. Bernardin, Konopaske, & Hagan, 2012; Levin, Siclair, Veniegas, & Taylor, 2002; Mount, Sytsma, Hazucha, & Holt, 1997; Newman, Hanges, & Outtz, 2007; Pulakos, White, Oppler, & Borman, 1989; Roth et al., 2003), comparisons between migrants with oriental background respectively their descendants and domestic Europeans are often drawn in Europe (Hanges & Feinberg, 2010). With the advent of technology the field of personnel selection has also changed. A recently suggested method is the application of video-based interviews in selection. Potential discrimination within this new method due to gender or ethnicity is still not sufficiently researched, especially in a European context. Therefore, the present study was undertaken to explore whether ethnicity or gender cause discrimination in ratings of video-based interviews.

Prior research on adverse impact in Europe

The results of research conducted in the Netherlands demonstrate substantive evidence for discrimination towards applicants of Arab origin (e.g. Derous, Ryan, & Nguyen, 2012;

Hiemstra, Derous, Serlie, & Born, 2012; Derous, Nguyen, & Ryan, 2009). For instance, Derous and Ryan (2012) illustrated that the rejection rate was 4.86 times higher for applicants bearing an Arab name than for Dutch. The “highly Dutch” applicants with a typical Dutch name and a strong Dutch affiliation in their résumés received the lowest rejection rate. The “highly Arab” group obtained the highest rejection rate. Furthermore, Derous and colleagues (2012) ascertained in an experimental design that implicit prejudices might trigger higher rejection rates against Arab minorities. Effects of adverse impact against ethnic minorities can be moderated by the motivation to control one’s own explicit prejudices (Deraus et al., 2012).

Discrimination against ethnic minorities in Germany

In spring 2014 results of a study conducted by the “Sachverständigenrat deutscher Stiftungen für Migration und Integration“ (SVR) (Expert Council of German Foundations on Migration and Integration), a Council focusing on the impact of migration and integration in Germany, documented the inert disadvantage that applicants of foreign origin are still suffering. More than 3,500 applications of a native German or an applicant of Turkish background – representing the biggest minority in Germany (Federal Statistical Office of Germany, 2013) – were sent to employers for two different apprenticeships training positions, namely motor mechanics and clerks. The net discrimination, the amount of which the “applicant” with a migration background received less positive resonance compared to the German native “applicant”, ranged between 3% for clerks and 17% for motor mechanics. While Germans on average needed to send in 5 applications in order to get at least one positive feedback, applicants with Turkish names¹ had to write more than 7. A substantive discrimination towards Turkish migrants and their descendants in the German labor market therefore became apparent (Forschungsbereich beim SVR, 2014). Whereas this research demonstrated that discrimination was primarily found in small companies, another German

¹ The applicants’ ethnic background was operationalized just via name, even the photo in the résumé was randomized.

Discrimination in video-based job interviews

study concerning discrimination in traineeships exhibited that discriminative effects disappeared when supplementary information was available in letters of reference (Kaas & Manger, 2010). Apart from these results little is known about discrimination within the German labor market with regard to ethnic minorities; until now no German field experiment exists that addresses the discrimination against applicants of different ethnicities. Thus, lab studies potentially provide valuable first insights into this research area. Additionally, the current study focuses on a selection step which occurs after a pre-selection based on written applications. A likely assessment instrument after pre-selection is an interview, given a large number of applicants a video-based interview is feasible. Therefore, the current study can be seen as the continuation of the work conducted by the Expert Council of German Foundations on Migration and Integration (2014). That study focused on the first step in a selection process, the written application. The current experiment now focuses on an interview taking place after this first selection.

Technology and fairness

Since 2006 the “Allgemeines Gleichbehandlungsgesetz (AGG)”, Germany’s General Equal Treatment Act (GETA), strengthens the rights of employees and applicants in Germany. In case of uncertainty employers have the burden of proof that the assessment tool is fair toward minorities (AGG, 2006). For example, in 2010 justice was administered in favor of a man from Ivory Coast by the labor court of Hamburg. He charged an organization for being unfair in their selection process because a telephone interview was held to preselect applicants. According to the court it is considered to be evident that a short contact by telephone is not adequate to judge the communication skills of an applicant. Furthermore, the judge declared that it was significantly more difficult for foreigners than for German native speakers to pass in a telephone interview (Antidiskriminierungsstelle des Bundes, 2014). Hence, it is important for Human Resource (HR) departments to use assessment techniques at

any stage of the selection process, which do not disadvantage applicants or employees with regard to their ethnic background, national origin, gender etc. (AGG, 2006).

Video-based assessment interviews

Structured interviews are on the one hand able to reduce adverse impact in selection processes (Hattrup & Roberts, 2010; Murphy, 2010; Ployhart & Holtz, 2008), but on the other hand might be inadequate for early stages of a multistage selection process because of the costs- financially and staff-wise- to develop and conduct an interview (Schmidt & Hunter, 1998). Internet-based interviews represent new instruments in HR selection processes to facilitate a fair and valid measurement and at the same time to reduce costs, while HR departments are enhancing international careers (Hiemstra et al. 2012). In a video-based and time-delayed interview neither synchronization in time nor in space is needed (Becker & Lindemann, 2012). Its format is similar to a structured interview, in which the questions and their order are pre-defined. When the respective HR department sends an interview invitation, applicants can choose the point in time to start the interview. Questions are depicted in written words, the answers are recorded on video after a pre-defined preparation time. The HR staff is able to watch these videos at a later date and give an evaluation using behaviorally anchored rating scales. Such tools are expected to allow for a valid pre-selection of applicants in an economic way. For recommending the use of this assessment instrument in Germany it is necessary to show in fairness with regard to ethnicity, sociocultural background, and gender, i.e. tests must not discover discrimination against any subgroup (Kubinger & Proyer, 2005; Reimann, Frenzel, Michalke, & Peper, 2008). While a video-based interview offers a lot more information than a resumé and discriminatory effects can disappear when more information is available (Kaas & Manger, 2010), it remains an empirical question whether discrimination against applicants of foreign origin will emerge.

Equal treatment for men and women is still a major topic on the political agenda when gender differences are discussed. The Federal Statistical Office of Germany estimates a “Gender Pay Gap” of 22%, i.e. women receive 22% less money than men (Federal Statistical Office of Germany, 2014). To some extent the difference can be ascribed to the fact that women work less frequently in leading positions or in occupations with better payment. Beside these facts, there is still one third of the “Gender Pay Gap” left, which could not be explained (Federal Statistical Office of Germany, 2014). In research about discrimination of women two concepts in relation to ethnic background emerged (e.g. Deros & Ryan, 2012). The first concept states that being a woman and part of an ethnic minority at the same time means to be part of two discriminated entities. This *double jeopardy hypothesis* (Browne & Misra, 2003) contrasts the *subordinate male target hypothesis* (Sidanius & Pratto, 1999), which implies that men from an ethnic minority are perceived as more aggressive than women of the same ethnic group. Therefore, men from an ethnic minority would receive the highest amount of discrimination (Bendick, Jackson, Reinoso, & Hodges, 1991). Because gender discrimination is a huge subject both in politics and academic research, this study also was concerned with the question whether there would be any discrimination of female and male applicants in ratings of video-based interviews. Gender effects are prominent whenever leading positions (Schein, Mueller, Lituchy, & Liu, 1996) or typically male occupations are concerned (Frauendorfer & Schmid Mast, 2013). Both do not apply to the job waiter/waitress chosen in this experiment. Additionally, due to the competing two concepts combining gender and ethnic background just outlined, a nondirectional hypothesis was phrased.

Agreeableness in appraisal processes

Research on social approaches suggests that the evaluation of skills is partly influenced by the rater’s own characteristics (Eiser & van der Pligt, 1984; Murphy Cleveland, Skatterbo, & Kinney, 2004). Personality traits can moderate the prejudice against certain subgroups. Besides conscientiousness the relation of a rater’s agreeableness and prejudiced appraisal

effects is well documented (e.g. Ekehammer & Akrami, 2003; Ekehammer, Akrami, Gylje, & Zakrisson, 2004; Bernardin, Cooke, & Villanova, 2000; Bernardin, Tyler, & Villanova, 2009). Agreeableness is associated with characteristics like humility, leniency, altruism, confidence and the ability of cooperation (Costa & Mc Crae; 1989). Bernardin and colleagues (2009) showed that people who score high on agreeableness scales and low on conscientiousness scales deliver a milder judgment in case of failure in performance. Because of the given evidence, we explored the relation of the raters' agreeableness and his or her given evaluation in order to test the robustness of possible discrimination effects. Of special interest are the conditions of rather inferior performances, when a milder judgment might occur (Bernardin et al. 2009; Yun et al., 2005).

Social Identity

To rule out a further alternative explanation for possible discrimination effects, the influence of inter-group phenomena was explored. According to the Social Identity Theory (Tajfel & Turner, 1986) every human being belongs to different social groups, which are supposed to be the foundation of social identity. In order to protect the social identity everyone tends to give a better evaluation of one's own group, the in-group, than of other social groups, the out-groups (Ashorth & Mael, 1989; Brewer & Gardner, 1996; Hestone & Jaspers, 1984; Turner, 1984; van Knippenberg, 1984). Ethnicity is supposed to be an important social group and as a part of one's own social identity an important influencing factor on how one evaluates the performance of others (Luhtanen & Crocker, 1992). It is therefore investigated here whether a discrimination effect can be explained partly by the migration background of the raters. Evidence not in line with the general findings demonstrated that there are conditions under which people tend to evaluate the out-group members better than their own in-group members. Brown (2000) explained that not only the social group determines the evaluation, but also the social status of the group. While members of a high-status group try to protect their higher social status, low-status group members

might count themselves to the high status in-group in order to change the social group and to enhance their social identity. If borders between the groups are fairly impermeable, low-status group members might focus on different skills in which their own in-group trumps the out-group (Brown, 2000). All in all, when testing discrimination within a selection instrument in a selection context, the social group of the assessor in terms of ethnicity should be regarded as a potential confounding variable. Therefore, the current study used ethnicity of the raters as control variable.

Aims of the study

Based on the preceding theoretical arguments pertaining to discrimination and the medium of video-based interviews, the following specific hypotheses were made:

H1. Discrimination effects will appear between German native applicants and applicants of Turkish origin.

H1a. A discrimination effect for communication skills will appear between German native applicants and applicants of Turkish origin.

H1b. A discrimination effect for team-working skills will appear between German native applicants and applicants of Turkish origin.

H2. A discrimination effect will appear between female and male applicants.

H3. The discrimination effects are moderated.

H3a. The raters' agreeableness moderates the discrimination effects between applicants (especially in the condition of rather inferior performance).

H3b. The raters' own migration background moderates the discrimination effects between German native applicants and applicants with Turkish migration background.

Methods

Design and procedure

The research questions were explored in a two (migration background – native German) by two (male – female) by two (higher quality – lower quality) factorial, experimental design. Four semi-professional actors, two men and two women, three of them from a Berlin drama school, one a hobby actor, mimed the applicants for a waiter/waitress position in an imaginary hotel chain. The job waiter represents a well-known job everybody immediately understands and has prior direct experience with (as a customer). Because participants were assumed to have no experience in HR management, we chose a job that could be visualized easily. Additionally, the operationalization of the waiter's required skills, communicational skill and teamwork, are highly understandable. Thus, the choice of the job waiter was meant to ensure that ratings could be given by inexperienced raters. Two of the actors had a Turkish background. Because it is well-known that language skills are confounding evaluations during assessment situations (Deprez-Sims & Morris, 2013; Hosoda & Stone-Romero, 2010) we made sure that none of the actors spoke with an accent. Thus, participants could recognize the foreign background only when the name of the “applicant” was given in the first interview answer (“Julian Lange” and “Julia Lange” vs. “Emin Yilmaz” and “Emine Yilmaz”) and by considering the physical appearance of the “applicants” (dark hair, tainted skin vs. light skin, blond hair). In order to avoid confounding effects of religion as much as possible we watched carefully that the actors did not wear any religious symbols, like crucifix, headscarf etc. Moreover, all actors wore white shirts. The circumstances were held as constant as possible; all “applicants” were of average appearance, all were wearing white shirts, the display window was held constant up to the chest. The “applicants” answers were scripted; additionally, the nonverbal expressions were kept similar between the candidates. All videos were recorded with the camera of the same notebook in the same room in order to hold exterior circumstances similar and to accommodate the real world assessment situation for an

applicant. Two skills, which are supposed to be essential for a waiter, were assessed in four interview questions with behaviorally anchored items. The first and the last interview question referred to communication skills and the second and third question to the capacity to work in teams:

1. Could you please introduce yourself and go into detail about formative situations in your vita.
2. In every workplace conflicts sometimes arise. Please illustrate a conflict situation which occurred in your past work and go into detail about how you did behave and how the situation came to an end.
3. Imagine a bus arrives - unexpectedly - with 60 people who wish to eat dinner in the restaurant. You are in charge of the indoor area this day, which is occupied by one third with 20 guests. Because of the nice weather the bus group chooses seats in the outdoor area, which your colleague is looking after. How do you react?
4. At the end of your video interview we invite you to conclude in describing what worked out well and what did not during this video interview.

All raters were invited to a lab. Before the video interview they were supposed to rate started, participants were introduced to the technology and workings of video-based interviews. Moreover, they were instructed about their role as an “HR manager” in the imaginative “restaurant chain”. After reading one of the interview questions, the participants started the pre-recorded video answer of the applicant. Raters were randomly assigned to one of the eight applicant videos. Literature suggests using behavioral anchored checklists rather than graphically anchored rating scales (Landy & Farr, 1980; Yun, Donahue, Dudley, & McFarland, 2005). The items of the behaviorally anchored scales followed the video on the computer and were assessed on a Likert scale in four steps; 1 = “it does not apply at all” – 2 = “it applies little” – 3 = “it applies well” – 4 = “it totally applies”. The means and standard

deviations of all eight cells are listed below (Table 1). After the video interview some descriptive data were collected. This was followed by the short-form of the Big Five Inventory (Rammstedt & John, 2005) which every participant filled out. Agreeableness was assessed with four items using a five-point Likert scale. Cronbach's alpha for Agreeableness in the current sample was .69.

Participants

Each participant evaluated one “applicant” in one out of eight conditions. 26 raters were randomly assigned to each of the conditions, altogether 208 data sets were collected. One condition was rated by 14 men and 12 women; every other condition was rated by the same number of men and women. Participants were recruited via a server of the Humboldt Universität zu Berlin. On average 35 minutes were needed to pass through the experiment. Four data sets were excluded because of problems in comprehension, technical disturbance or test objectivity. The age of the raters varied between 18 and 75 years ($M = 32.87$, $SD = 13.56$). In the descriptive questions 38 of the participants affirmed to have a migration background. While 76.4% of them did not have any experience with observation-based techniques, 3 participants indicated to have acquired high expertise in HR assessment processes in their occupational tasks. Regarding their occupation, about 60% of the participants stated to be in training. Most of the participants were students (56%). Nearly 20% were in employment, while about 8% were retired. Distributions of age, experience, migration background and occupational mode are fairly similar in the eight conditions.

Statistical analyses

Hypothesis 1: The effects of three independent variables (migration background – gender – quality of the answer) and their interactions were estimated on four dependent variables - referring to the four interview questions and their evaluation - in a multivariate analyses of variance (MANOVA). A main effect of “migration background” represents

support for the first hypothesis that a discrimination effect between German native applicants and applicants of Turkish origin emerged. In subsequent univariate ANOVAs effects for each dependent variable were tested. Thus, it will be possible to test, if there is an overall discrimination or whether discrimination is unique to specific interview questions.

Hypothesis 2: The main effect of “gender” will clarify the question, if there is any discrimination regarding male applicants or female applicants. Furthermore, effects of the univariate ANOVAs will test, if discrimination is unique to specific interview questions.

Hypothesis 3: In a multivariate analyses of covariance (MANCOVA) the raters’ “agreeableness” and their migration status were entered as covariates. If any of the covariates reached statistical significance or if the results of the MANOVA were to change considerably, evidence would have been found that discrimination effects were moderated by agreeableness or the rater’s migration status.

Results

The Pillar-trace statistics of the multivariate analyses of variance (MANOVA) was used because the box test reached significance, $F(70.00, 54763.26) = 2, p < .01$ (Bray & Maxwell, 1985, p. 34). In the global test a moderate and significant effect for the applicants’ migration background was found, $F(4, 197) = 1.64, p < .05, \eta^2 = .06$. Holding everything else constant, applicants with Turkish migration background reached a better evaluation than German natives. No significant gender effect occurred. The large and significant main effect for the quality of the applicants’ answers demonstrates that the experimental manipulation worked well, $F(4, 197) = 228.16, p < .01, \eta^2 = .82$.

In subsequent univariate ANOVAs it became apparent that the effect for migration background was statistically significant for both scales measuring communication skills and for one scale measuring capacity for teamwork (3rd interview question). For the communication skills assessed in the last interview question interactions needed to be

considered. Both two-way interactions, “gender x migration background” and “gender x answer’s quality” are best understood in the three-way interaction, $F(4, 197) = 4.39, p < .01, \eta^2 = .08$ (see Table 1

Means and standard deviations grouped by dependent variables.

		German				Turkish			
		Male		Female		Male		Female	
		M	SD	M	SD	M	SD	M	SD
Communication1	rather superior	13.58	1.36	13.81	1.96	14.50	1.45	14.23	1.37
	rather inferior	10.62	1.70	9.19	1.65	10.73	2.01	10.62	1.92
Communication2	rather superior	13.85	1.59	14.08	2.00	14.77	1.37	15.04	1.08
	rather inferior	11.12	2.34	8.58	1.90	10.23	2.29	10.65	1.94
Teamwork1	rather superior	13.69	1.76	13.77	1.63	14.50	1.33	14.42	1.30
	rather inferior	7.31	1.49	6.92	2.15	7.00	2.33	7.46	2.34
Teamwork2	rather superior	13.38	2.08	12.77	2.20	13.58	1.92	14.38	1.53
	rather inferior	6.08	1.60	6.92	2.10	6.58	1.98	6.92	2.94

Table). It turned out that while the appraisal of the applicants in this condition was quite similar in the rather superior answering format condition, the German female applicant was obviously devaluated in the rather inferior answering format condition. In the remaining dependent variables, main effects can be interpreted as such because no hierarchical differences appeared in interactions (

Figure 1).

A MANCOVA with the covariates “agreeableness” and “migration background” (see Table 2) of the raters yielded a marginally significant, moderate effect for the raters’ own migration status, $F(4, 195) = 2.10, p < .08, \eta^2 = .04$. No significant effect emerged for “agreeableness” as covariate, $F(4, 195) = .51, p = .73, \eta^2 = .01$. Both covariates did not reach statistical significance. Neither considering the personality trait² nor the raters’ own migration background altered the effects of the independent variables and their interactions (Table 1

Means and standard deviations grouped by dependent variables.

		German				Turkish			
		Male		Female		Male		Female	
		M	SD	M	SD	M	SD	M	SD
Communication1	rather superior	13.58	1.36	13.81	1.96	14.50	1.45	14.23	1.37
	rather inferior	10.62	1.70	9.19	1.65	10.73	2.01	10.62	1.92
Communication2	rather superior	13.85	1.59	14.08	2.00	14.77	1.37	15.04	1.08
	rather inferior	11.12	2.34	8.58	1.90	10.23	2.29	10.65	1.94
Teamwork1	rather superior	13.69	1.76	13.77	1.63	14.50	1.33	14.42	1.30
	rather inferior	7.31	1.49	6.92	2.15	7.00	2.33	7.46	2.34
Teamwork2	rather superior	13.38	2.08	12.77	2.20	13.58	1.92	14.38	1.53
	rather inferior	6.08	1.60	6.92	2.10	6.58	1.98	6.92	2.94

² Conscientiousness ($M = 3.75$; $SD = .68$; *Cronbach's* $\alpha = .67$) as a covariate yielded no significant effect, $F(4/196) = .75, p < .56, \eta^2 = .02$. It did not change the effects reported here either.

Table). Therefore, we did not find any evidence for moderating effects.

Discussion

In an experimental design we explored the influence of ethnicity and gender of applicants on possible discrimination in ratings of video-based interviews. The results of our sample showed that we did not find any evidence for discrimination against women or against the ethnic minority. Furthermore, applicants of Turkish origin received better evaluations compared to German natives. This – at first sight – contra-intuitive result proved to be robust when the covariates “agreeableness” and the raters’ own “migration background” were also considered.

Effect of ethnicity in performance ratings

As postulated in the Hypothesis 1, a discrimination effect, i.e. a different evaluation between German native applicants and applicants of Turkish origin, occurred. However, the applicants with a migration background received better evaluations in the skills that were investigated, “communication skills” and “capacity for teamwork”. Holding everything else constant, applicants of Turkish origin achieved a better rating for their answers in the video-based assessment interview than the German native applicants did. Subsequent ANOVAs showed that in three of the four dependent variables (DV) this result was statistically significant. These results were only partially explained by interaction effects. While in one DV, “communication2”, non-ordinal two-way and three-way interactions occurred, the remaining effects represent real main effects and may be interpreted as such (Bühner & Ziegler, p. 418f.).

This result might be confusing at first because former studies of “adverse impact” have found evidence for discrimination of the ethnic minority rather than a preference for the minority (see above). But research related to the “Social Identity Theory” (Tajfel & Turner, 1986) provides some evidence for the current finding. It was shown that people of a lower

status group sometimes receive a higher evaluation of both, their in-group members and of the higher status out-group members (Brown, 2000). On the one hand the in-group members (of the lower status group) are trying to enhance their social identity by giving a higher appraisal to their own in-group members. In contrast, it illustrates an act of protection of the higher status, when high status group members favor the low status out-group members (Brown, 1984 p. 608-609; van Knippenberg, 1984, p. 572).

Effect of gender on performance ratings

Results of the MANOVA yielded no significant effect for the applicants' gender. In our sample it did not matter which gender an applicant had for the evaluation in the investigated skills: Neither male nor female applicants received systematically higher ratings for their answers. Moreover, the significant interactions of "migration background" x "gender" and "gender" x "quality of answer" in the dependent variable "Communication2" should be seen as a result of the significant three-way interaction. Hence, no evidence for the *double jeopardy hypothesis* (Browne & Misra, 2003) or the *subordinate male target hypothesis* (Sidanius & Pratto, 1999) was provided. Thus, the hypothesis, that a gender discrimination occurred, was rejected. This does not necessarily mean that video-based assessments are immune to gender discrimination. A more likely explanation could be the choice of the job here. Prior research has shown that discrimination of female applicants was more likely when hiring for a leading position (Schein, Mueller, Lituchy, & Liu, 1996) or for typical male occupations (Frauendorfer & Schmid Mast, 2013). Moreover, Ziegler, Dietl, Danay, Vogel, and Bühner (2011) were able to show that no gender discrimination occurred in structured and unstructured interviews designed to select German applicants for an apprenticeship. Thus, future research should systematically vary those aspects in order to replicate the here reported gender fairness.

Interactions for the dependent variable “Communication2”

As mentioned before, significant interactions in the dependent variable “Communication2” occurred. The two-fold interactions “gender x migration” and “gender x quality of answer” reached significance in the omnibus test only. The univariate analyses only showed significant interactions for the second communication skill variable. The same finding occurred for the three-way interaction. Taking a closer look at the kind of interaction, it turned out, that both two-way interactions could be merged in the three-way interaction; in the rather inferior format the German female applicant got a strikingly lower evaluation in the last interview question (related to the 2nd communication variable “Communication2”) compared with all other applicants. So, it is the female German applicant, who was devaluated in this specific condition. It is unclear, where this devaluation originates. The female German actress represented the role model for all other actors. Thus, a chance effect cannot be ruled out. Any further interpretations should be based on replications of this effect.

No moderating effects of the investigated covariates

In order to test the robustness of the effects and to establish substantial moderators two covariates were tested. A small to moderate but not significant effect was revealed for the covariate “participants’ own migration background”. Potentially the rater’s ethnicity mattered, when applicants of different origin were evaluated. Because “raters’ own migration background” failed to reach significance as a covariate, this interpretation must be handled with caution, even more so because this marginally significant finding only occurred for the “Communication2” variable. Research on the “Social Identity Theory” (Tajfel & Turner, 1986) nevertheless supports this finding; ethnicity is an important dimension for building the social identity. When rater and ratee are belonging to different ethnicities and therefore to different social groups, differences in evaluation of performance might manifest.

The second covariate “agreeableness” did not influence the performance rating of applicants at all. In neither condition any moderating effect emerged. The performance rating of video-based assessment interviews under the conditions operationalized here did not change according to the amount of the rater’s agreeableness.

It turned out quite clearly that the effects were robust against included covariates. The distinction of raters with and without migration background as well as the consideration of their personality trait “agreeableness” did not change effects from the MANOVA in significance or effect size.

Limitations

It is a valid limitation, that the social groups of raters with and without a migration background were not a clear entity. In the descriptive question it was merely asked, if the rater had a migration background. Neither nationality nor origin were recorded. It is highly questionable, which social group is taking an effect on evaluation of the applicants, in case the rater is neither of German nor of Turkish origin. On the other hand, a larger moderating effect seems possible, when raters and applicants are of the same ethnicity. In this sample, the social group of the raters with migration background was probably fairly heterogeneous in contrast to the group of raters without any migration background. Further investigation can answer the question, if more homogenous social groups related to the ethnicity yield bigger moderating effects. Additionally, the lack of an experimental control for the operationalization of ethnicity as part of the social identity of the raters might limit the interpretation of the effects as a discrimination against certain ethnic groups. In order to hold the video interview as realistic as possible ethnic group membership was not made especially salient in the experiment. Thus, it is conceivable that there are individual differences in how central ethnicity was for the raters’ social identity.

More research is needed to clarify, whether the higher evaluation for applicants of foreign origin is a chance finding only. Besides evidence related to the “Social Identity Theory” (Taijfel & Turner, 1986) we do not know of any other research with the result, that ratees of the minority were preferred to ratees of the majority. Even though these effects showed up quite robustly, an interpretation should be taken with caution, only, until further supporting evidence is provided.

We must emphasize that the design of a lab study is limiting the generalizability of the findings. In order to disclose discriminating practices against job applicants of different gender or ethnicity a field study should be the logical next step.

Despite these limitations, the current study provides first evidence that video-based assessments can be utilized within selection contexts without having to fear discrimination effects due to gender or ethnicity.

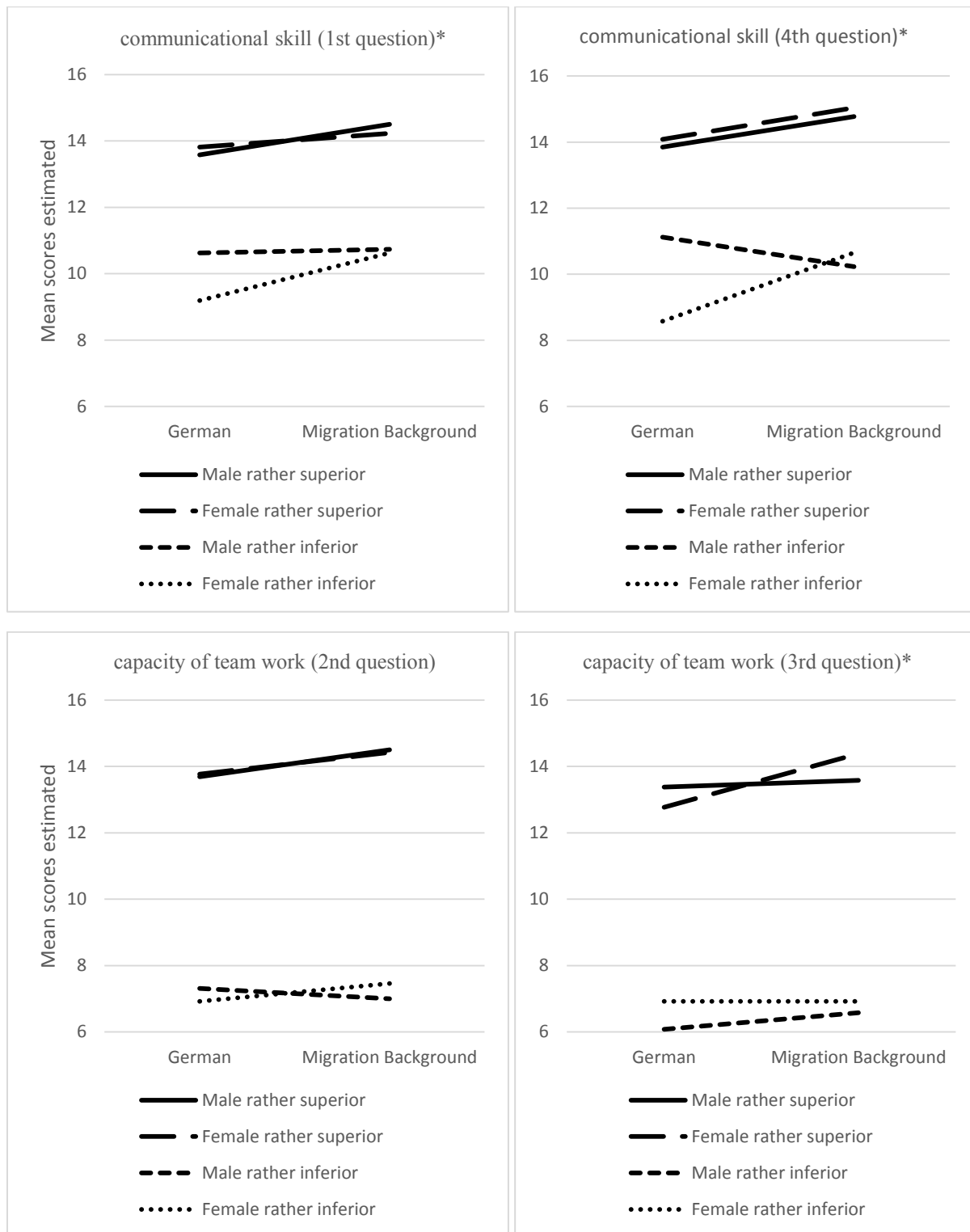


Figure 1. Mean scores were estimated by subsequent ANOVAs representing the mean evaluation of the applicant. * Mean differences of German native applicants and applicants of Turkish origin at a level of significance $p < .05$.

Table 1

Means and standard deviations grouped by dependent variables.

		German				Turkish			
		Male		Female		Male		Female	
		M	SD	M	SD	M	SD	M	SD
Communication1	rather superior	13.58	1.36	13.81	1.96	14.50	1.45	14.23	1.37
	rather inferior	10.62	1.70	9.19	1.65	10.73	2.01	10.62	1.92
Communication2	rather superior	13.85	1.59	14.08	2.00	14.77	1.37	15.04	1.08
	rather inferior	11.12	2.34	8.58	1.90	10.23	2.29	10.65	1.94
Teamwork1	rather superior	13.69	1.76	13.77	1.63	14.50	1.33	14.42	1.30
	rather inferior	7.31	1.49	6.92	2.15	7.00	2.33	7.46	2.34
Teamwork2	rather superior	13.38	2.08	12.77	2.20	13.58	1.92	14.38	1.53
	rather inferior	6.08	1.60	6.92	2.10	6.58	1.98	6.92	2.94

Table 2

Results of the multivariate analysis of covariance (MANCOVA) and subsequent ANOVAs.

<i>Effect / Source</i>	<i>DV</i>	<i>F</i>	<i>df_{effect}</i>	<i>df_{res}</i>	<i>p</i>	<i>Partial Eta²</i>	<i>Power</i>
<i>Agreeableness</i>	<i>MANOVA</i>	.51	4	195	.73	.01	.17
	<i>Communication1</i>	.03	1		.87	<.001	.05
	<i>Communication2</i>	1.44	1		.23	.01	.22
	<i>Teamwork1</i>	.27	1		.61	<.001	.08
	<i>Teamwork2</i>	.75	1		.39	<.001	.14
<i>Participants' migration background</i>	<i>MANOVA</i>	2.10	4	195	.08	.04	.62
	<i>Communication1</i>	.39	1		.53	<.001	.10
	<i>Communication2</i>	3.19	1		.08	.02	.43
	<i>Teamwork1</i>	.25	1		.61	<.001	.08
	<i>Teamwork2</i>	.10	1		.75	<.001	.06
<i>Migration background</i>	<i>MANOVA</i>	3.00	4	195	.02*	.06	.79
	<i>Communication1</i>	9.56	1		<.01**	.05	.87
	<i>Communication2</i>	7.55	1		.01*	.04	.78
	<i>Teamwork1</i>	2.80	1		.10	.01	.38
	<i>Teamwork2</i>	3.64	1		.06	.02	.48
<i>Gender</i>	<i>MANOVA</i>	1.55	4	195	.19	.03	.47
	<i>Communication1</i>	3.00	1		.08	.01	.41
	<i>Communication2</i>	2.05	1		.15	.01	.30
	<i>Teamwork1</i>	.00	1		1.00	<.001	.05
	<i>Teamwork2</i>	1.41	1		.24	.01	.22
<i>Quality of answer</i>	<i>MANOVA</i>	225.05	4	195	<.01**	.82	1.00
	<i>Communication1</i>	250.72	1		<.01**	.56	1.00

Migration background x Gender	Communication2	272.39	1		<.01**	.58	1.00
	Teamwork1	729.67	1		<.01**	.79	1.00
	Teamwork2	561.92	1		<.01**	.74	1.00
	MANOVA	2.09	4	195	.08	.04	.62
Migration background x Quality of answer	Communication1	.81	1		.37	<.001	.15
	Communication2	7.64	1		.01*	.04	.79
	Teamwork1	.50	1		.48	<.001	.11
	Teamwork2	.57	1		.45	<.001	.12
Gender x Quality of answer	MANOVA	.54	4	195	.71	.01	.18
	Communication1	.03	1		.86	<.001	.05
	Communication2	.29	1		.59	<.001	.08
	Teamwork1	1.46	1		.23	.01	.23
Migration background x Gender x Quality of answer	Teamwork2	1.17	1		.28	.01	.19
	MANOVA	2.64	4	195	.04*	.05	.73
	Communication1	2.43	1		.12	.01	.34
	Communication2	7.10	1		.01*	.03	.76
Migration background x Gender x Quality of answer	Teamwork1	.01	1		.93	<.001	.05
	Teamwork2	.68	1		.41	<.001	.13
	MANOVA	4.54	4	195	<.01**	.09	.94
	Communication1	3.47	1		.06	.02	.46
Migration background x Gender x Quality of answer	Communication2	8.99	1		<.01**	.04	.85

Discrimination in video-based job interviews

<i>Teamwork1</i>	.89	1	.35	<.001	.16
<i>Teamwork2</i>	2.61	1	.11	.01	.36

Notes: Arranged according to the covariates, “agreeableness” and “participants’ migration background”, the independent variables “migration background”, “gender”, “quality of answer” and their interactions and the dependent variables; *represents statistical significance at $p < .05$; **represents statistical significance at $p < .01$.

References

Allgemeines Gleichbehandlungsgesetz (AGG) vom. (2006). (BGBl. I S. 1897), das zuletzt durch Artikel 8 des Gesetzes vom, 2013 (BGBl. I S. 610) geändert worden ist.

Available at: <http://www.gesetze-im-internet.de/agg/index.html>.

Antidiskriminierungsstelle des Bundes (2014). *Ausgewählte Entscheidungen deutscher Gerichte zum Antidiskriminierungsrecht*. Retrieved from http://www.antidiskriminierungsstelle.de/SharedDocs/Downloads/DE/publikationen/Rechtsprechungs%C3%BCbersicht/rechtsprechungsuebersicht_zum_antidiskriminierungsrecht.pdf?__blob=publicationFile

Ashforth, B. E., & Mael, F. (1989). Social Identity Theory and the Organization. *The Academy of Management Review*, 14, 20–39.

Becker, M., & Lindemann, S. (2012). Auswahl ohne Vorurteile. In *E-Recruiting* (pp. 20–22). Retrieved from http://www.viasto.com/wp-content/uploads/2012/11/201211_Zeitversetzte-Videointerviews_Auswahl-ohne-Vorurteile.pdf

Bendick, M. Jr., Jackson, C., Reinoso, V., & Hodges, L. (1991). Discrimination against Latino job applicants: A controlled experiment. *Human Resource Management*, 30, 469-484. Doi: 10.1002/hrm.3930300404.

Bernardin, H. J., Cooke, D. K., & Villanova, P. (2000). Conscientiousness and agreeableness as predictors of rating leniency. *Journal of Applied Psychology*, 85, 232–236.

Bernardin, H.J., Konopaske, R., & Hagan, C. M. (2012). A comparison of adverse impact levels based on top-down, multisource, and assessment center data: Promoting diversity and reducing legal challenges. *Human Resource Management*, 51(3), 313-341.

- Bernardin, H. J., Tyler, C. L., & Villanova, P. (2009). Rating level and accuracy as a function of rater personality. *International Journal of Selection and Assessment*, 17, 300–310.
- Bray, J. H., & Maxwell, S. E. (1985). *Multivariate Analysis of Variance: Sage University Paper Series on Quantitative Applications in the Social Sciences*. Newbury Park, CA: Sage.
- Brewer, M. B., & Gardner, W. (1996). Who is this "We"? Levels of collective identity and selfrepresentations. *Journal of personality and social psychology*, 71, 83–93.
- Brown, Rupert J. (1984). The role of similarity in intergroup relations. In H. Tajfel (Ed.). *The Social Dimension, Volume 2. European developments in social psychology*. (pp. 603–623). Cambridge: Cambridge University Press.
- Brown, R. (2000). Social identity theory: Past achievements, current problems and future challenges. *European Journal of Social Psychology*, 30, 745–778.
- Browne, I. & Misra, J. (2003). The intersection of gender and race in the labor market. *Annual Review of Sociology*, 29, 487–513.
- Bühner, M., & Ziegler, M. (2009). *Statistik für Psychologen und Sozialwissenschaftler*. [Statistics for Psychologist and Social Scientists] München: Pearson Studium.
- Costa, P.T., Jr. & McCrae, R.R. (1989). *The NEO PI/FFI manual supplement*. Odessa, FL: Psychological Assessment Resources.
- Dean, M. A., Roth, P. L., & Bobko, P. (2008). Ethnic and gender subgroup differences in assessment center ratings: a meta-analysis. *Journal of Applied Psychology*, 93, 685–691.

Deprez-Sims, A. & Morris, S.B. (2013). The Effect of Non-native Accents on the Evaluation of Applicants during an Employment Interview: The development of a path model. *International Journal of Selection and Assessment*, 21, 355 – 367.

Derous, E., Nguyen, H.-H., & Ryan, A. M. (2009). Hiring discrimination against Arab minorities: Interactions between prejudice and job characteristics. *Human Performance*, 22, 297– 320.

Derous, E., & Ryan, A. M. (2012). Documenting the Adverse Impact of Résumé Screening: Degree of ethnic identification matters. *International Journal of Selection and Assessment*, 20, 464–474.

Derous, E., Ryan, A. M., & Nguyen, H. D. (2012). Multiple categorization in resume screening: Examining effects on hiring discrimination against Arab applicants in field and lab settings. *Journal of organizational behavior*, 33, 544–570.

Eiser, J. R., & van der Pligt, Joop. (1984). Attitudes in a social context. In H. Tajfel (Ed.), *The Social Dimension, Volume 2. European developments in social psychology* (pp. 363–378). Cambridge: Cambridge University Press.

Ekehammar, B., & Akrami, N. (2003). The relation between personality and prejudice: a variable-and a person-centred approach. *European Journal of Personality*, 17, 449–464.

Ekehammar, B., Akrami, N., Gylje, M., & Zakrisson, I. (2004). What matters most to prejudice: Big Five personality, Social Dominance Orientation, or Right-Wing Authoritarianism? *European Journal of Personality*, 18, 463–482.

Federal Statistical Office [Statistisches Bundesamt] (2013). *Datenreport 2013*. Wiesbaden: Statistisches Bundesamt. Available at

https://www.destatis.de/DE/Publikationen/Datenreport/Downloads/Datenreport2013.pdf?__blob=publicationFile.

Federal Statistical Office [Statistisches Bundesamt] (2014a). *Auf dem Weg zur Gleichstellung? Bildung, Arbeit und Soziales- Unterschiede zwischen Frauen und Männern*. Wiesbaden: Statistisches Bundesamt. Available at [https://www.destatis.de/DE/Publikationen/Thematisch/Bevoelkerung/Querschnitt/Weg zurGleichstellung5120001149004.pdf?__blob5publicationFile](https://www.destatis.de/DE/Publikationen/Thematisch/Bevoelkerung/Querschnitt/Weg_zurGleichstellung5120001149004.pdf?__blob5publicationFile).

Federal Statistical Office [Statistisches Bundesamt] (2014b). *Bevölkerung und Erwerbstätigkeit. Bevölkerung mit Migrationshintergrund. Ergebnisse des Mikrozensus 2013. (Fachserie 1 Reihe 2.2)*. Wiesbaden: Statistisches Bundesamt. Available at https://www.destatis.de/DE/Publikationen/Thematisch/Bevoelkerung/MigrationIntegration/Migrationshintergrund2010220137004.pdf?__blob5publicationFile.

Forschungsbereich beim Sachverständigenrat deutscher Stiftungen für Integration und Migration (2014). *Diskriminierung am Ausbildungsmarkt*. Ausmaß, Ursachen und Handlungsperspektiven. Berlin: Königsdruck Printmedien und digitale Dienste GmbH. Retrieved from http://www.svr-migration.de/wp-content/uploads/2014/11/SVR-FB_Diskriminierung-am-Ausbildungsmarkt.pdf

Frauendorfer, D. & Schmid Mast, M. (2013). Hiring Gender-Occupation Incongruent Applicants: The Positive Impact of Recruiter Interpersonal Sensitivity. *Journal of Personnel Psychology*, 12, 182-188.

Hanges, P. J., & Feinberg, E. G. (2010). International Perspectives on Adverse Impact: Europe and Beyond. In J. L. Outtz (Ed.), *Adverse impact: Implications for*

- organizational staffing and high stakes selection* (pp. 349–373). New York: Taylor & Francis Group.
- Hattrup, K., & Roberts, B. G. (2010). What Are the Criteria for Adverse Impact? In J. L. Outtz (Ed.), *Adverse impact: Implications for organizational staffing and high stakes selection* (pp. 161–197). New York: Taylor & Francis Group.
- Hewstone, M., & Jaspars, J. M. F. (1984). Social dimensions of attribution. In H. Tajfel (Ed.), *The Social Dimension, Volume 2. European developments in social psychology* (pp. 379–404). Cambridge: Cambridge University Press.
- Hiemstra, A. M. F., Derous, E., Serlie, A. W., & Born, M. P. (2012). Fairness perceptions of video resumes among ethnically diverse applicants. *International Journal of Selection and Assessment*, 20, 423–433.
- Hosoda, M. & Stone-Romero, E. (2010). The effect of foreign accents on employment-related decisions. *Journal of Managerial Psychology*, 25, 113-132.
- Kaas, L., & Manger, C. (2010). Ethnic discrimination in Germany's labour market: a field experiment. *German Economic Review*, 13, 1-20.
- Kraiger, K., & Ford, J. K. (1985). A meta-analysis of ratee race effects in performance ratings. *Journal of Applied Psychology*, 70, 56–65.
- Kubinger, K. D., & Proyer, R. (2005). Gütekriterien. [Quality criteria] In K. Westhoff, et al. (Eds.), *Grundwissen für die berufsbezogene Eignungsbeurteilung nach DIN 33430* [Basic knowledge for occupational assessment according to DIN 33430] (2nd ed., pp. 191–199). Lengerich: Pabst.
- Landy, F. J., & Farr, J. L. (1980). Performance rating. *Psychological Bulletin*, 87, 72–107.

- Levin, S., Sinclair, S., Veniegas, R. C., & Taylor, P. L. (2002). Perceived discrimination in the context of multiple group memberships. *Psychological Science, 13*, 557–560.
- Luhtanen, R., & Crocker, J. (1992). A collective self-esteem scale: Self-evaluation of one's social identity. *Personality and Social Psychology Bulletin, 18*, 302–318.
- Mount, M. K., Sytsma, M. R., Hazucha, J. F., & Holt, K. E. (1997). RATER-RATEE RACE EFFECTS IN DEVELOPMENTAL PERFORMANCE RATINGS OF MANAGERS. *Personnel Psychology, 50*, 51–69.
- Murphy, K. R. (2010). How a Broader Definition of the Criterion Domain Changes Our Thinking About Adverse Impact. In J. L. Outtz (Ed.), *Adverse impact: Implications for organizational staffing and high stakes selection* (pp. 137–160). New York: Taylor & Francis Group.
- Murphy, K. R., Cleveland, J. N., Skattebo, A. L., & Kinney, T. B. (2004). Raters Who Pursue Different Goals Give Different Ratings. *Journal of Applied Psychology, 89*, 158–164.
- Newman, D. A., Hanges, P. J., & Outtz, J. L. (2007). Racial groups and test fairness, considering history and construct validity. *American Psychologist, 26*, 1082–1083.
- Outtz, J. L., & Newman, D. A. (2010). A Theory of Adverse Impact. In J. L. Outtz (Ed.), *Adverse impact: Implications for organizational staffing and high stakes selection* (pp. 53-94). New York: Taylor & Francis Group.
- Ployhart, R. E., & Holtz, B. C. (2008). The diversity–validity dilemma: Strategies for reducing racioethnic and sex subgroup differences and adverse impact in selection. *Personnel Psychology, 61*, 153–172.
- Pulakos, E. D., White, L. A., Oppler, S. H., & Borman, W. C. (1989). Examination of race and sex effects on performance ratings. *Journal of Applied Psychology, 74*, 770–780.

- Rammstedt, B., & John, O. P. (2005). Kurzversion des Big Five Inventory (BFI-K)[Short version of the Big Five Inventory (BFI-K)]. *Diagnostica*, 51(4), 195–206.
- Reimann, G., Frenzel, T., Michalke, S., & Peper, M. (2008). Verbreitung und Akzeptanz der DIN 33430: Eine Stellungnahme [Distribution and acceptance of the DIN 33430: A comment]. *Zeitschrift für Personalpsychologie*, 7, 178–180.
- Roth, P. L., Huffcutt, A. I., & Bobko, P. (2003). Ethnic group differences in measures of job performance: a new meta-analysis. *Journal of Applied Psychology*, 88, 694–706.
- Sackett, P. R., & DuBois, C. L. (1991). Rater-ratee race effects on performance evaluation: Challenging meta-analytic conclusions. *Journal of Applied Psychology*, 76, 873–877.
- Schein, V., Mueller, R., Lituchy, T., & Liu, J. (1996). Think manager-think male: a global phenome-non? *Journal of Organizational Behavior*, 17, 33-41.
- Sidanius, J. & Pratto, F. (1999). Social dominance: An intergroup theory of social hierarchy and oppression. New York: Cambridge University Press.
- Tajfel, H., & Turner, J. C. (1986). The social identity theory of intergroup behavior. In S. Worchel & W. G. Austin (Eds.), *Psychology of intergroup relations* (2nd ed., pp. 7-24). Chicago: Nelson-Hall.
- Turner, J. C. (1984). Social identification and psychological group formation. In H. Tajfel (Ed.), *The Social Dimension, Volume 2. European developments in social psychology* (pp. 518– 539). Cambridge: Cambridge University Press.
- van Knippenberg, Ad F. M. (1984). Intergroup differences in group perceptions. In H. Tajfel (Ed.), *The Social Dimension, Volume 2. European developments in social psychology* (pp. 560–578). Cambridge: Cambridge University Press.

Yun, G. J., Donahue, L. M., Dudley, N. M., & McFarland, L. A. (2005). Rater personality, rating format, and social context: Implications for performance appraisal ratings.

International Journal of Selection and Assessment, 13, 97–107.

Ziegler, M., Dietl, E., Danay, E., Vogel, M. & Bühner, M. (2011). Predicting Training Success with General Mental Ability, Specific Ability Tests, and (Un)Structured Interviews: A meta-analysis with unique samples. *International Journal of Selection and Assessment*, 19, 170-182.