Received: 10 November 2022 | Accepted: 11 February 2024

DOI: 10.1111/apps.12528

ORIGINAL ARTICLE

APPLIED PSYCHOLOGY IAAP

Do ethnic, migration-based, and regional language varieties put applicants at a disadvantage? A meta-analysis of biases in personnel selection

Niklas Schulte[1] | **Johannes M. Basch**[2] | **Hannah-Sophie Hay**[3] | **Klaus G. Melchers**[3]

[1]Freie Universität Berlin, Berlin, Germany

[2]Hochschule Neu-Ulm, Neu-Ulm, Germany

[3]Universität Ulm, Ulm, Germany

Correspondence
Niklas Schulte, Freie Universität Berlin, Arbeitsbereich Psychologische Diagnostik, Differentielle und Persönlichkeitspsychologie, Habelschwerdter Allee 45, 14195 Berlin, Germany.
Email: n.schulte@fu-berlin.de

Abstract

This meta-analysis examined biases in personnel selection owing to applicants' use of non-standard language such as ethnic and migration-based language varieties or regional dialects. The analysis summarized the results of 22 studies with a total N of 3615 raters that compared applicants with an accent or dialect with applicants speaking standard language. The primary studies used different standard and non-standard languages and assessed different dependent variables related to hiring decisions in job interviews. The $k = 109$ effect sizes (Hedges' g) were assigned to the dependent variables of competence, warmth, and hirability. Non-standard speakers were rated as less competent ($\delta = -0.70$), less warm ($\delta = -0.17$), and less hirable ($\delta = -0.51$) compared to standard speakers. Thus, at the same level of competence, non-standard speakers are rated lower than standard speakers and might, therefore, be disadvantaged in personnel selection contexts. We also considered several potential moderator variables (e.g., applicants' specific language

The authors would like to thank Ella Voigtländer for her help with data coding.

variety, raters' own use of non-standard language, and raters' background) but only found rather limited support for them. Furthermore, publication bias had only limited effects. Practical implications for personnel selection are discussed.

KEYWORDS

accent, dialect, hirability, job interview, language, personnel selection

INTRODUCTION

Globalization and migration, which have been increasing for decades, have increased workforce mobility (Martins et al., 2004; McAuliffe & Triandafyllidiou, 2021). Many people are seeking work far from home. One thing that they bring with them is the language shaped by their origin: a language variety of the local language that is influenced by their mother tongue or another foreign language. Likewise, domestic employees may speak local dialects from other regions within a country. Previous research found that speaking these types of non-standard languages can impair interpersonal evaluations compared to when people speak in the standard language of a certain region (Fuertes et al., 2012).

When people speak with non-standard varieties such as accents or dialects in a high-stakes context, this might negatively affect important outcomes. Specifically, applicants who speak non-standard languages might face biases concerning their evaluations in employment interviews, assessment centers, or in/during other selection procedures, in which oral communication skills play a role. As a result, they may have a lesser chance of receiving a job offer.

Previous studies on the impact of non-standard language in personnel selection found heterogeneous results ranging from clearly negative (e.g., Michelsen, 2019) to neutral or even sometimes positive effects (e.g., Hosoda & Stone-Romero, 2010). Other studies suggest that non-standard speakers are perceived as less competent and hirable but as warmer (e.g., Roessel et al., 2019). Furthermore, it is unclear whether certain variables such as the raters' own language-related background moderate the impact of speaking non-standard language on how applicants are evaluated.

Given these open questions, we conducted a meta-analysis to systematically investigate the effects of non-standard language in personnel selection. Building on the stereotype content model (Fiske et al., 2002), we organized the different outcomes investigated in the primary studies in two groups of interpersonal evaluations, namely, competence and warmth. Additionally, we analyzed the effects on perceived hirability. To further explain the heterogeneity of effects, we considered several potential moderators. By doing so, we were able to draw a more nuanced picture of potentially biasing effects based on non-standard language in personnel selection and to develop an agenda for further research.

STANDARD VERSUS NON-STANDARD LANGUAGE AND POTENTIAL EFFECTS ON INTERPERSONAL EVALUATIONS

Within a language, a distinction can be made between standard language and different non-standard language varieties. In linguistics, standard language is defined as the language that is used in the mass media and in the educational system of a country (Yule, 2006). It is also referred to as the supra-regional standard variety, written language, or high variety and is established and maintained by norm authorities such as teachers, linguists, and language editors of publishing houses (Ammon, 2004). Thus, the standard language is the version of a language that is taught to someone learning this language as a foreign language. Furthermore, it is accepted as the official language of a certain country. Therefore, although a certain dialect (e.g., Scottish dialect) might be spoken by the majority of people in a certain region (e.g., the Highlands), Standard British English would still be considered as the standard language of Scotland. Any variation of this standard language is called non-standard language (Yule, 2006).

There are different types of non-standard language such as accents and dialects. Accents are deviations from the standard language in pronunciation and accentuation (e.g., someone speaking English with a French accent; Giles, 1970). A further distinction can be made, on the one hand, between people who are native speakers but who speak a language variety that is characteristic of a particular region within the area where their mother tongue is spoken (native regional; e.g., southern US accent). On the other hand, people can speak a language variety that deviates from the standard language based on their or their parents' country of origin (native ethnic; e.g., somebody who grew up in a family of non-native speakers).

Dialects are language varieties spoken in specific regions that are a smaller part of the area where a certain language is referred to as the standard language (e.g., the Bavarian dialect in Germany; Mai & Hoffmann, 2010). People speaking with a dialect may also have an accent in the standard language—but dialect goes beyond an accent, as it may also include specific pronunciation, vocabulary, and grammar (Yule, 2006). In the following, we will use the term non-standard language to summarize both accents and dialects.

Similar to features such as sex, skin color, or age, linguistic variations are also noticed early in social interactions (Fiske & Neuberg, 1990) and therefore influence how people classify others into social categories (Tajfel & Wilkes, 1963). These categories, in turn, are linked to psychological traits attributed to speakers based on their accent or dialect (Lambert et al., 1960). The information that is quickly retrieved through this stereotyping can be useful during interactions and can provide orientation and reassurance to the conversation partners (Fiske, 2004). However, rapid information processing based on stereotypes can lead to discrimination if negative characteristics are attributed to the speakers, possibly resulting in negatively biased treatment of the speakers (Tajfel, 1970).

Personnel selection is one of the areas in which the attribution of negative characteristics based on the use of non-standard language becomes particularly relevant. If applicants are evaluated more negatively because of speaking non-standard language, this can lead to negative consequences, such as a lower probability of receiving a job offer or being promoted. However, previous findings show a heterogeneous pattern of the effects of (non-)standard language in personnel selection regarding the direction and the size of the effects (Michelsen, 2019; Sterk, 2020).

PREVIOUS RESEARCH AND DEVELOPMENT OF HYPOTHESES

There are two main theories that are often used to explain the influence of non-standard language on evaluations of applicants in personnel selection. The first is accent prestige theory (Giles, 1970), which has been used in earlier research on language perception and focuses on the variables' status and solidarity. Status refers to the perceived social prestige, education, or success of a speaker, whereas solidarity refers to the perceived trustworthiness, similarity, and friendliness of a person (Giles, 1970; see also Bradac & Wisegarver, 1984, for a later extension of accent prestige theory).

Research related to accent prestige theory has usually found that the status of standard speakers is higher than that of individuals with non-standard language varieties (e.g., Cheyne, 1970; Giles & Billings, 2004; Giles & Sassoon, 1983). In line with this, a meta-analysis from Fuertes et al. (2012) on the effects of non-standard language on interpersonal evaluations found that people who speak without an accent are ascribed higher social status ($d = 0.99$) and higher solidarity ($d = 0.52$). Thus, speaking non-standard language can lead to the stereotype of having a lower status and may, therefore, result in potential disadvantages for people speaking non-standard language when other negative attributes of a lower status (e.g., lower formal education) are activated (Fuertes et al., 2012). Similarly, the lower solidarity ratings for people speaking non-standard language might be paralleled by lower affection toward them.

In light of the meta-analytic results from Fuertes et al. (2012), a relevant question is whether the use of non-standard language also impairs applicants' chances in selection situations. In contrast to accent prestige theory, perceptions of social status, for example, are only of indirect relevance in a selection context, as interviewers and recruiters do not have to evaluate applicants' social status but their competence and hirability. Competence is one of the two dimensions of social perception from the stereotype content model (Fiske et al., 2002) that is often used in more recent research on the effects of non-standard language. However, in some previous studies that used separate measures for competence and prestige, high correlations between the two constructs were found (Fiske et al., 2002, Study 3). Furthermore, in other studies, competence was used as an indicator of prestige (e.g., Bradac & Wisegarver, 1984). Thus, there is a potential overlap between the prestige dimension from accent prestige theory and the competence dimension from the stereotype content model.

In addition to competence, the other main dimension of the stereotype content model is warmth, which can be described as the degree of affection toward another person (Fiske et al., 2002). It has some similarities with solidarity, which refers to the perceived kindness and friendliness of a person. However, as the stereotype content model is more directly relevant for personnel selection and therefore more common in this field, we decided to use it as the theoretical framework for our meta-analysis.

In contrast to accent prestige theory, the stereotype content model was developed not only for the linguistic context but also to explain the effects of interpersonal impression formation and of stereotypes in general. In line with this, it has also often been used in recent studies from organizational settings (Cuddy et al., 2012, for an overview) dealing with the perceptions of applicants (e.g., Amaral et al., 2019). In contrast to language-related characteristics that we investigate in the present meta-analysis, other studies from the personnel selection domain focusing on the stereotype content model often dealt with application documents in order to investigate differences in perceived competence and warmth. As an example, a Swedish study

by Agerström et al. (2012) found that in order to get invited equally often to job interviews, Arab applicants have to appear comparably warmer and more competent than Swedish applicants.

As noted above, the first dimension of the stereotype content model is competence. When individuals are perceived as competent, they are seen as intelligent, capable, and skillful, and they are also given more respect (Fiske et al., 2002). Several studies from different countries investigating different accents or dialects found that people who speak the standard language are usually considered to be more intelligent and more competent and thus more suitable (see, for example, De la Zerda & Hopper, 1979, for the effects of a Mexican accent in comparison to standard American English; Giles et al., 1981, for the effects of a Welsh accent in comparison to a British accent; and Rakić et al., 2011, for the effects of different German dialects in comparison to standard German). Based on the stereotype content model and on previous findings, we therefore predict the following:

> **Hypothesis 1.** Non-standard-language-speaking applicants are rated lower than standard-speaking applicants regarding their competence.

The second dimension of the stereotype content model is warmth. As noted above, it can be described as the amount of affection toward another person. Perceptions of warmth are relevant for personnel selection because if people are considered warm, they are also seen as agreeable, trustworthy, and easy to get along with (Fiske et al., 2002). All these qualities make it easier to work together. Further, these qualities are associated with the personality trait of agreeableness, which in turn has been shown to predict interpersonal teamwork behavior (see also Tasa et al., 2011) and team performance (Lim et al., 2023). In line with this, several studies (e.g., Fetscherin et al., 2020; Varghese et al., 2018) found that higher perceptions of applicants' warmth and competence are associated with higher hirability. In contrast to perceptions of competence, some studies found that people speaking non-standard language are rated higher on warmth. Giles et al. (1981), for example, found that people who spoke with a Welsh accent were perceived as more honest, kind-hearted, good-natured, generous, and agreeable while appearing also less aggressive and status-related. Similarly, Bradac and Wisegarver (1984) found that English speakers with Mexican accents were rated as more supportive compared to US Americans. Based on these results, we predict the following:

> **Hypothesis 2.** Non-standard-language-speaking applicants are rated higher than standard-speaking applicants regarding their warmth.

In a personnel selection context, perceptions of competence and warmth are potential influencing factors when it comes to evaluations of applicants' hirability or promotability (e.g., Amaral et al., 2019). Specifically, higher perceptions of competence should lead to higher perceptions of applicants' hirability. Furthermore, in a selection context, it seems likely that perceptions of competence are more important than perceptions of warmth, so non-standard language should negatively impact hirability or promotability ratings. In line with this, some studies that directly assessed hirability (or promotability) not only found effects on competence and warmth but also found that non-native speakers were less likely to be nominated for managerial positions than native speakers (Huang et al., 2013) and that speaking a regional German dialect impaired applicants' chances to be invited for another interview compared to standard

German (Bald & Kanning, 2019). Similarly, Timming (2017) and Rakić et al. (2011) found that applicants who spoke non-standard language were less likely to be rated suitable for a job, especially for customer-facing jobs (Timming, 2017). Based on these results, we predict:

> **Hypothesis 3.** Non-standard-language-speaking applicants are less likely to be hired than standard-speaking applicants.

Potential moderators of the effects of standard versus non-standard language

Apart from the direct effects of standard versus non-standard language, there are some variables that might moderate the aforementioned relationships. One of these moderators is whether raters themselves speak a non-standard language variety. Several conceptual arguments make it likely that evaluations will be more positive in such a situation. First, such an effect can be explained by the mere-exposure effect (Zajonc, 1968): If people are used to non-standard language through their own accent or dialect, they might perceive non-standard language in others as more pleasant and sympathetic. Furthermore, the similarity–attraction effect (Byrne, 1971) might be relevant, because other people are evaluated more positively if they are perceived as more similar with regard to relevant factors such as race, gender, ethnicity, socio-economic background, or age.

Considering the role of similarity in workplace interactions, similarity can lead to more successful and effective interaction between individuals on the one hand and to discrimination, and prejudice against people who are unlike their colleagues, supervisors, or subordinates on the other hand (Abbasi et al., 2023). In line with this, a recent review by Abbasi et al. (2023) supported the assumptions from similarity–attraction theory for many workplace interactions such as recruitment and selection. More specifically, ethnic similarity, for example, was found to be related to interview ratings in campus interviews in a study by Goldberg (2005). Furthermore, a study by Chen and Lin (2014) found that the perceived person–recruiter fit (recruiters' perceptions of how well applicants' personality matches their own personality) predicted hiring recommendations. Furthermore, the review also pointed out that other variables such as organizational measures or legislation to protect disadvantaged groups might override or moderate the effects of similarity.

In addition to the mere-exposure and the similarity–attraction effect, social identity theory (Tajfel, 1974) states that people attach cognitive and affective importance to their group membership and link their self-worth to it. To maintain self-worth, they evaluate others who belong to the same social group more favorably and people from other groups less favorably. Finally, blatant prejudice can also affect evaluations of others (Horverak et al., 2013). In line with this, meta-analytic evidence supports the occurrence of bias against minority groups in personnel selection contexts such as older applicants (Batinovic et al., 2023), Muslims (Bartkoski et al., 2018), or based on physical appearance or ethnicity (Lippens et al., 2023). Accordingly, for selection situations, in which verbal interactions occur (e.g., interviews), applicants speaking non-standard language might trigger prejudices against them and these prejudices might be stronger in raters speaking standard language.

Taken together, according to all the different conceptual explanations, it seems reasonable that speaking in a non-standard language oneself also leads to more benevolent (or less negative) evaluations of other non-standard speaking individuals. However, direct research testing this prediction in personnel selection is sparse and inconsistent to date (Bald & Kanning, 2019). Nevertheless, there is some research that supports this view at least indirectly: In a study examining brain activity, for example, Hernández et al. (2019) found that hearing one's own language variety in others activates the reward network and triggers a sense of belonging. Similarly, raters also perceived voices more positively that matched their own language variety (Giles, 1971). Because of the inconclusive findings regarding the effect of the raters' own dialect, we raise the following research questions:

Research Question 1: Is the effect of speaking non-standard language on perceived competence moderated by the raters' use of non-standard language?

Research Question 2: Is the effect of speaking non-standard language on perceived warmth moderated by the raters' use of non-standard language?

Furthermore, there is also some evidence that the beneficial effects of raters speaking a language variety themselves might not only affect ratings of competence and warmth. For example, Bald and Kanning (2019) found that applicants with a Bavarian dialect, who were evaluated by raters speaking with a variety themselves, were more likely to be invited for a second interview after an initial telephone interview and less likely to receive a direct rejection than their standard German-speaking peers. However, the findings concerning this effect are very sparse and inconsistent and, therefore, do not allow us to make a clear prediction. Accordingly, this leads to the following research question:

Research Question 3: Is the effect of speaking non-standard language on perceived hirability moderated by the raters' use of non-standard language?

Apart from the raters' own use of non-standard language, we also wanted to exploratively investigate further possible moderators that could potentially have an impact on applicants' evaluations. These moderators included the type of raters who served as participants (HR professionals vs. students), raters' mean age, and applicants' specific language variety. The type of raters might be relevant, as raters with more expertise sometimes provide less bias (e.g., Marlowe et al., 1996) and more valid ratings (Gaugler et al., 1987; Lievens, 2001). In addition, we investigated the specific language variety to find out whether there are differences between applicants speaking with a foreign accent from a particular country (e.g., a Mexican accent while speaking English), a specific ethnic variety (e.g., African American English), or a specific regional dialect (e.g., Bavarian German).

We also evaluated whether the experimental design moderated the effects of non-standard language. Experiments on the effect of language varieties on hiring decisions usually use multigroup experimental designs in which participants evaluate applicants' behaviors, abilities, or characteristics. The language variety of the applicants is manipulated (i.e., there is one standard language and at least one non-standard language condition). The verbal statements (e.g., the job interview responses) of the rating targets (i.e., the alleged applicants) are typically prerecorded with the same text in standard and non-standard language. There are two different techniques for presenting the different conditions: the matched-guise technique and the verbal-guise technique. Matched-guise means that the same individual is speaking standard language in one condition and non-standard language in the other condition and verbal-guise means that different individuals are speaking standard language in one condition and non-standard language in the other. The matched-guise technique (e.g., Duroska et al., 2014; Hosoda et al., 2012; Michelsen, 2019) controls better for speaker characteristics that might additionally influence

ratings (e.g., physical attractiveness or speaking speed) and the verbal-guise technique (e.g., Deprez-Sims & Morris, 2013; Hosoda & Stone-Romero, 2010; Timming, 2017) ensures the most authentic language use possible by using speakers who are perfectly familiar with each language variety. Our moderator analyses tested whether this design choice (i.e., matched-guise vs. verbal-guise) influences studies' effect sizes.

METHOD

Literature search

We searched the electronic databases PsycINFO, Web of Science, and Google Scholar. In addition to English search terms, we also used a German search string for the database PSYNDEX, which covers German psychological literature. The search strings linked a part for the linguistic component of the meta-analysis with a part for the personnel selection context. The terms for the linguistic component were as follows: accent*, dialect*, non-standard speech*, non-standard language*, regional language*, language variet*, pronunciation*, non-native speaker*, migra* language*, or immigra* language. The terms for the personnel selection context were: job interview*, assessment cent*, hire*, hiring, applica*, recruiting, personnel selection*, employee selection*, employment interview*, job candidate*, job suitab*, employment suitab*, employability, or selection interview*. At least one word from each of the two parts (language and personnel selection) had to appear in the title, abstract, or content of the article. In Google Scholar, the search for the words was limited to the title of the article. The words from both parts could be linked together in any way. For the complete list of search strings on all databases, see Appendix A in the online supplemental materials at https://osf.io/zrv73/. The search in the databases was completed on May 27, 2021. In addition to the database search results, 13 articles were identified through a backward search of the reference lists of the identified studies. In sum, the literature search resulted in 680 articles of which 56 were removed as duplicates, resulting in 624 unique publications (see Figure 1).

Inclusion and exclusion criteria

We screened the title, abstract, and, when necessary, the full text of each article. We included studies that compared spoken non-standard language with a supra-regional standard language in a job application context (e.g., a selection interview). The target studies used ratings of applicants' competence, warmth, and/or hirability. We included field and laboratory studies, regardless of the country in which they were conducted. Qualitative studies and studies that did not examine human speakers (e.g., computer voices) were excluded. Other exclusion criteria were language disorders, sociolect (language used by a socioeconomic class, profession, age group, etc.), and studies that investigated low language registers (e.g., low vocabulary) or individual language habits (e.g., idiolects). The studies were selected by the third author. If there was any doubt about the inclusion or exclusion of a specific study, the other authors were consulted, and a joint decision was made. Together, 19 studies (containing 23 samples) were suitable for inclusion (see Figure 1).

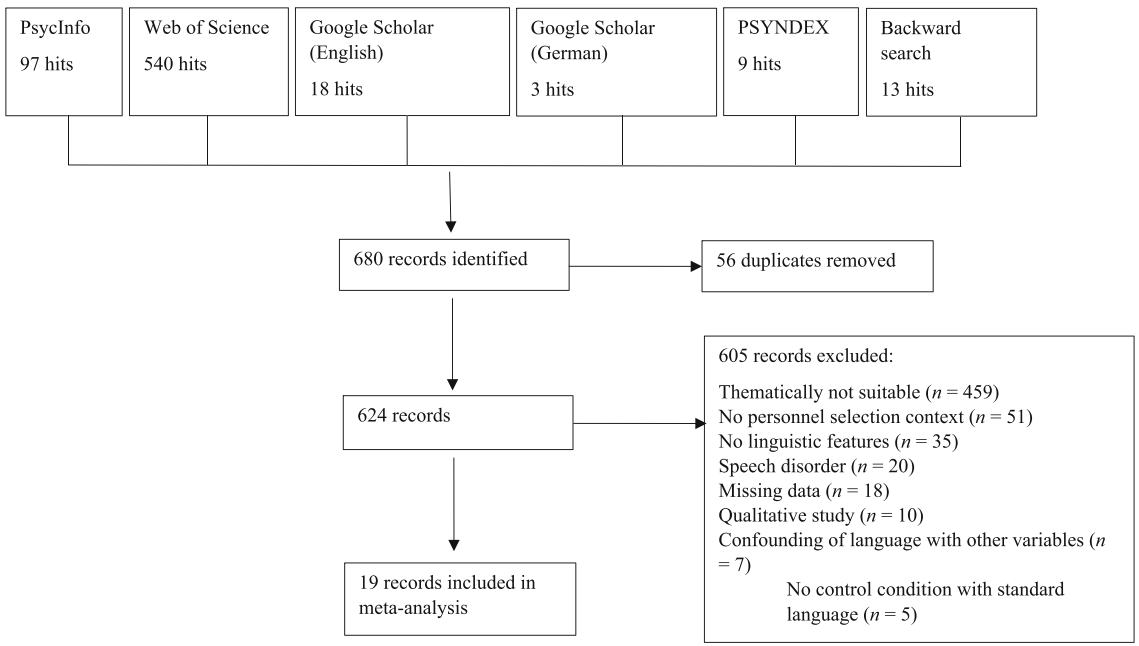

FIGURE 1 Flowchart for selecting the primary studies to be included in the meta-analysis.

Underlying study designs

All identified primary studies were laboratory studies that used experimental designs with one or more alleged applicants speaking the standard language and one or more non-standard varieties of this standard language, using either the matched- or verbal-guise technique. Furthermore, all primary studies used either Standard US/British English or German as the standard language. Although we wanted to consider field studies from real applied settings as well, none of the studies found in the literature search used such a sample.

Coding

The coding process included data on five variable types: information about the publication, the study methodology, the applicants, the raters, and the effect sizes. On the publication level, we coded the publication year and the publication type (published paper, Bachelor/Master thesis, or dissertation). On the study level, we coded the continent, country, and federal state of data collection, sample size, the standard and non-standard language varieties covered in the study, independent variables (conditions as labeled in the primary study), whether the study was conducted online or in a laboratory, study type (experiment, quasi-experiment, or correlational study), and study design (between- or within-subjects design). Moreover, we coded the selection methods used (e.g., telephone interview or video interview sometimes in combination with a resume and photo of the applicant), the medium that was used for the presentation of interview responses (audio or video), the accent guise (matched or verbal), and any potential covariates considered in the primary study.

With respect to the alleged applicants, we coded their non-standard language variety type (native ethnic, such as African American English; native regional, such as the German dialect Bavarian, or non-native such as English with a Dutch accent), the specific language variety, their mean age, the percentage of female applicants, and the sector for which an applicant had to be selected.

Regarding the raters, we coded the rater type (human resources managers, students, or heterogeneous convenience sample), their relation to the study's standard language (native, non-native, or mixed), their mother tongue or original language variety, their mean age, the percentage of female raters, the percentage of raters with a migration background, and the percentage of raters with experience in personnel selection. Additionally, we coded matches of applicants' and raters' language variety and whether the raters spoke any non-standard variety themselves. Finally, the means and standard deviations of the ratings for the experimental groups were used to calculate effect sizes, and the reliabilities (Cronbach's αs) of the dependent variables were also coded.

The dependent variables were first coded with their original names from the respective primary studies. In the second step, the dependent variables were assigned to one of three categories, namely, competence, warmth, and hirability. The category system was determined before coding. The assignment of the dependent variables to the three categories was conducted by three coders holding a PhD in Psychology who fully agreed with each other in their assignments. The coding manual and the coded data are available via https://osf.io/zrv73/.

One of the authors then coded all the studies that were included in the meta-analysis and a second person coded 9 of the 19 articles so that we were able to determine coder agreement.

For the coded effect sizes, we used the Pearson correlation as the indicator of coder reliability and found perfect agreement, $r = 1.00$. For the other coded variables, the mean coder agreement was 90%.

Effect sizes and meta-analytic approach

First, we calculated standardized mean differences (Cohen's *ds)* as effect sizes. The mean value of the dependent variable in the control condition (i.e., the standard language) was subtracted from the mean value of the language variety. The result was divided by the standard deviation of the control condition (Smith & Glass, 1977). Thus, positive effect sizes indicate a positive effect of non-standard language. As Cohen's *d* has a slight bias in small samples (i.e., it overestimates the true standardized mean difference), Cohen's *d* was converted to Hedges' *g*, as Hedges' *g* is an unbiased estimator (Borenstein et al., 2009; even though the differences typically concern at most the third decimal place and the interpretation therefore remains unchanged, Rosenthal, 1994). Hedges' *g*s were then corrected for the unreliability of the measurement of the dependent variable in the respective study. For artifact corrections, missing reliabilities were replaced by the mean Cronbach's α of the respective criterion categories (competence: $\alpha = .89$, warmth: $\alpha = .82$, hirability: $\alpha = .91$).

Analyses were conducted using the *metafor* package (Version 3.0–2; Viechtbauer, 2010) in R (R Core Team, 2021). The *Q*-test for heterogeneity was significant for all three dependent variables: $Q(27) = 1384.77$ for competence, $Q(17) = 108.17$ for warmth, and $Q(62) = 2402.75$ for hirability, all $ps < .001$. In light of the significant *Q*-test results and due to theoretical considerations, we estimated a random effects model using restricted maximum likelihood estimation.

Moderator analyses

To reduce the risk of false positive significance test results in our moderator analyses, we considered only those variables that were theoretically suitable and practically relevant and that were reported in an appropriate number of primary studies. For this purpose, we determined the number of effect sizes per moderator or their manifestations and selected the following variables for moderator analyses: applicants' language variety type, applicants' specific language variety, rater type, raters' relation to the study's standard variety, accent guise, the match of applicants' and raters' language variety, whether the raters spoke any non-standard variety, raters' mean age, the percentage of raters being migrants, and the percentage of raters with experience in personnel selection. For tables reporting the absolute frequencies of moderator levels, see Appendix B for categorical and Appendix C for continuous moderators in the online supplemental materials.

Publication bias

Meta-analytic results can be biased, for instance, because studies finding non-significant results are less likely to be published. To investigate a potential publication bias, we inspected the funnel plot graphically. Here, we paid particular attention to whether studies with small sample sizes and above-average effect sizes were overrepresented. In addition, we tested for funnel plot

asymmetry using Egger's test (Egger et al., 1997) in combination with the trim and fill method (Duval & Tweedie, 2000), which provides a meta-analytic effect size estimate corrected for potentially existing but unpublished studies.

RESULTS

In sum, 19 articles reporting 23 studies with a total of 3615 participants were included in the present meta-analysis. The analyses were based on $k = 109$ effect sizes (28 on competence, 18 on warmth, and 63 on hirability). Thus, some studies reported more than one effect size, as they examined different dependent variables or multiple experimental conditions (e.g., different language varieties). The number of effect sizes (k), the sample size N, and the meta-analytic effect sizes uncorrected (g) and corrected for unreliability (δ) are shown in Table 1.

Our hypotheses assumed that applicants speaking non-standard language are perceived as less competent (Hypothesis 1), but warmer (Hypothesis 2), and are finally less likely to be hired (Hypothesis 3). As can be seen in Table 1, the meta-analytic effect sizes were negative for all dependent variables. According to Cohen's (1992) conventions for standardized mean differences, the effect size for competence was large, $g = -.66$, $\delta = -0.70$, $p_\delta < .001$, whereas it was small for warmth, $g = -.15$, $\delta = -0.17$, $p_\delta = .04$, and moderate for hirability, $g = -.48$, $\delta = -0.51$, $p_\delta = <.001$. Thus, overall, applicants speaking non-standard language were perceived as considerably less competent, a little colder, and less hirable than applicants speaking standard language. These results support Hypotheses 1 and 3 (on competence and hirability) but provide no support for Hypothesis 2 (on warmth).

The forest plots show the distribution of effect sizes in the primary studies (see Appendix D of the online supplement). Of the 28 effect sizes on competence, the vast majority (i.e., 26) were negative (see Figure D1 of the online supplement). Thus, the available evidence is rather clear and consistent for competence. In contrast, the results of the primary studies for warmth were more equivocal (see Figure D2). Although the overall estimate based on the 18 effect sizes was

TABLE 1 Results of the meta-analysis of the relationship between language variety and competence, warmth, and hirability.

Dependent variable	k	N	g	δ	SE_δ	p_δ	95% CI_ρ		80% CR_ρ	
							LB	UB	LB	UB
Competence	28	2140	−0.66	−0.70	0.12	<.001	−0.94	−0.46	−0.86	−0.54
Warmth	18	1407	−0.15	−0.17	0.09	.04	−0.34	−0.01	−0.29	−0.06
Hirability	63	3334	−0.48	−0.51	0.07	<.001	−0.65	−0.37	−0.60	−0.42
Across all dependent variables	109	3615	−0.48	−0.51	0.06	< .001	−0.62	−0.40	−0.58	−0.43

Note: k = number of effect sizes; N = sample size; g = uncorrected estimate for standardized mean difference; δ = estimate of Hedges' g corrected for unreliability of the dependent variable; SE_δ = standard error of the corrected estimate of Hedges' g; p = p-value of the corrected estimate of Hedges' g; 95% CI_δ = 95% confidence interval of the estimate of Hedges' g corrected for the unreliability of the criterion; 80% CR_δ = 80% credibility interval of the corrected estimate of Hedges' g. All values are based on a random effects model.

Abbreviations: LB = lower bound of the interval; UB = upper bound of the interval.

negative, a third (i.e., six) of the included primary study effect sizes were positive, with four of these positive effect sizes being from a single publication by Roessel et al. (2019). Finally, the primary study results on hirability were somewhat more homogeneous (see Figure D3). Of the 63 effect sizes, 50 were negative. Nevertheless, the results on all three dependent variables showed some heterogeneity indicating that third variables may moderate the effects examined.

Moderator analyses

To identify relevant boundary conditions of the described effects, we conducted moderator analyses for the three dependent variables (competence, warmth, and hirability) separately. For each of them, we investigated seven categorical variables as potential moderators and conducted *Q*-tests to see whether each potential moderator would explain heterogeneity in effect sizes. The categorical moderators were applicants' language variety type, applicants' specific language variety, rater type (heterogeneous convenience sample, human resources managers, and students), raters' relation to the study's standard variety (native, non-native, or mixed), the accent guise (matched- vs. verbal-guise), the match of applicants' and raters' language variety, and whether the raters themselves spoke any non-standard variety. We report the direct parameter estimates for the respective moderator expressions in Tables 2–4 separately for the dependent variable competence, warmth, and hirability.[1]

Moderator analyses of the relationship between language varieties and competence

Table 2 shows the results for moderators of competence ratings. As can be seen there, we did not find effect size differences on the aggregated level of applicants' language variety type (native ethnic, native regional, or non-native), $Q_M(2) = 1.48$, $p = .48$. However, on a more fine-grained level, differences between applicants' specific language variety emerged, $Q_M(12) = 43.96$, $p < .001$. For example, the effects were particularly strong for the German dialect Saxon ($\delta = -1.45$, $p < .001$) and for Hong Kong English ($\delta = -1.90$, $p < .001$). Thus, applicants speaking these varieties were particularly disadvantaged by others in the assessment of their competence. However, the effect size estimates for individual dialects should be interpreted with caution, as some of them were based on a single primary study only. Furthermore, there was no significant difference between the mean effect sizes for the different groups of raters, $Q_M(2) = 0.23$, $p = .89$. Likewise, we did not find differences between effect sizes for raters' relation to the study's standard variety, $Q_M(2) = 0.50$, $p = .78$, or for the accent guises that the different studies used to present the language varieties, $Q_M(1) = 1.46$, $p = .23$, even though the mean effect size from studies using the verbal-guise technique was descriptively larger than from studies using the matched-guise technique. Finally, Research Question 1 asked whether effects varied depending on whether applicants and raters both spoke the same dialect. The match of applicants' and raters language variety did not moderate the effect sizes, $Q_M(2) = 0.06$, $p = .97$. Furthermore, the effect did not depend on whether the raters spoke any dialect themselves, $Q_M(2) = 1.76$, $p = .42$.

TABLE 2 Analysis of the categorical moderators of the relationship between language variety and competence.

Moderator variables	*k*	R^2	Hedges' *g*	δ	SE_δ	*p*	95% CI_δ LB	95% CI_δ UB
Applicants' language variety type	28	0%						
Native ethnic	7		−0.61	−0.66	0.25	.009	−1.15	−0.16
Native regional	15		−0.58	−0.61	0.17	<.001	−0.93	−0.28
Non-native	6		−0.92	−0.99	0.27	<.001	−1.51	−0.46
Applicants' specific language variety	28	56%						
Standard language: Standard English								
AAVE	2		−0.51	−0.57	0.31	.06	−1.17	0.03
Standard language: American English								
African American English	4		−0.60	−0.64	0.24	<.01	−1.12	−0.17
Hong Kong	1		−1.75	−1.90	0.47	<.001	−2.83	−0.97
Mexican	1		−0.82	−0.88	0.43	.04	−1.73	−0.04
Strong German accent	3		−0.88	−0.94	0.25	<.001	−1.43	−0.44
Weak German accent	1		−0.38	−0.40	0.44	.36	−1.26	0.46
Standard language: Standard German								
Bavarian	5		−0.19	−0.20	0.19	.30	−0.58	0.18
Berlin	1		−0.45	−0.49	0.45	.27	−1.36	0.39
Berlin and Saxon and Bavarian and Cologne	1		−0.71	−0.76	0.44	.08	−1.62	0.10
Saarland	1		−0.34	−0.38	0.44	.39	−1.25	0.49
Saxon	5		−1.40	−1.45	0.19	<.001	−1.83	−1.07
Swabian	2		0.37	0.41	0.31	.19	−0.19	1.01
Turkish	1		−0.84	−0.87	0.43	.04	−1.72	−0.02
Rater type	28	0%						
Heterogeneous convenience sample	13		−0.63	−0.65	0.18	<.001	−1.01	−0.29
Human resources managers	4		−0.60	−0.64	0.35	.07	−1.33	0.04
Students	11		−0.72	−0.78	0.20	<.001	−1.17	−0.38
Raters' relation to the study's standard variety	16	0%						
Native	5		−0.66	−0.69	0.29	.02	−1.26	−0.13
Non-native	4		−0.75	−0.80	0.31	<.01	−1.41	−0.19
Mixed	7		−0.49	−0.53	0.24	.02	−1.00	−0.07
Accent guise	28	2%						
Matched-guise	11		−0.48	−0.51	0.20	<.01	−0.90	−0.13
Verbal-guise	17		−0.77	−0.82	0.16	<.001	−1.12	−0.51
	21	0%						

(Continues)

TABLE 2 (Continued)

Moderator variables	k	R^2	Hedges' g	δ	SE_δ	p	95% CI_δ LB	95% CI_δ UB
Match of applicants' and raters' language variety								
No	6		−0.71	−0.75	0.32	.02	−1.39	−0.12
Yes	2		−0.60	−0.64	0.55	.24	−1.72	0.43
Mixed	13		−0.64	−0.66	0.21	<.01	−1.08	−0.24
Any non-standard variety rater	20	0%						
No	4		−0.79	−0.81	0.38	.03	−1.55	−0.06
Yes	9		−0.86	−0.91	0.26	<.001	−1.41	−0.41
Mixed	7		−0.37	−0.41	0.29	.16	−0.97	0.16

Note: All analyses are based on a random effects model; k = number of effect sizes included in the analyses; R^2 = variance explained in the heterogeneity of effect sizes by the model with significant moderators; Hedges' g = uncorrected estimate of Hedges' g; δ = estimate of Hedges' g corrected for criterion unreliability; SE_δ = standard error of the corrected estimate of Hedges' g; p = p-value of the corrected estimate of Hedges' g for the respective factor level; 95% CI_δ = 95% confidence interval of the corrected estimate of Hedges' g.
Abbreviations: AAVE = African American Vernacular English; LB = lower bound of the interval; UB = upper bound of the interval.

Moderator analyses of the relationship between language varieties and warmth

Table 3 shows the results for moderators of warmth ratings. As for competence, warmth ratings did not depend on applicants' language variety type, $Q_M(2) = 5.66$, $p = .06$), but on their specific language variety, $Q_M(10) = 104.94$, $p < .001$. For instance, applicants who spoke African American English or who spoke British English with a moderate Dutch accent were rated less warm than standard speakers in the respective studies, while other varieties showed positive effects (but those were usually covered only in single studies; for effect sizes, see Table 3).

In contrast to the analyses for competence, two additional moderator variables were significant for ratings of warmth. First, the rater type was a significant moderator, $Q_M(2) = 26.81$, $p < .001$, and the mean effect was negative when raters were human resources managers ($\delta = -0.72$, $p < .001$) but weakly positive in student raters ($\delta = 0.17$, $p = .07$). Second, concerning Research Question 2, the raters' relation to the study's standard variety significantly influenced the warmth ratings, $Q_M(2) = 9.03$, $p = .01$. Specifically, raters rated only the applicants speaking with non-standard language as significantly less warm than standard speakers when the standard variety of the respective study was their mother tongue ($\delta = -0.50$, $p = <.001$) but not if the raters' mother tongue was different from the standard variety of the respective study ($\delta = -0.04$, $p = .68$) or when the sample was mixed in this respect ($\delta = 0.08$, $p = .71$). In contrast to this, neither the accent guise used in the respective study, $Q_M(1) = 0.035$, $p = .85$, nor the match of applicants' and raters' language varieties, $Q_M(2) = 5.40$, $p = .07$, nor the presence of any accent or dialect in the raters, $Q_M(1) = 0.84$, $p = .36$, significantly influenced the assessment of warmth.

TABLE 3 Analysis of the categorical moderators of the relationship between language variety and warmth.

Moderator	k	R^2	Hedges' g	δ	SE_δ	p	95% KI_δ LB	95% KI_δ UB
Applicants' language variety type	18	14%						
Native ethnic	5		−0.44	−0.50	0.17	.003	−0.83	−0.18
Native regional	2		0.07	0.08	0.23	.74	−0.38	0.53
Non-native	11		−0.08	−0.10	0.10	.35	−0.29	0.10
Applicants' specific language variety	18	100%						
Standard language: American English								
African American English	4		−0.65	−0.72	0.12	<.001	−0.95	−0.48
German	2		−0.39	−0.45	0.09	<.001	−0.64	−0.27
Mexican	1		0.09	0.09	0.10	.36	−0.10	0.29
Spanish	2		−0.10	−0.12	0.09	.21	−0.30	0.07
Strong German accent	3		0.18	0.19	0.07	<.01	0.05	0.33
Weak German accent	1		0.29	0.31	0.12	<.01	0.08	0.55
Standard language: British English								
Moderate Dutch accent	1		−0.44	−0.48	0.10	<.001	−0.68	−0.28
Slight Dutch accent	1		−0.20	−0.23	0.10	.03	−0.43	−0.02
Standard language: Standard German								
Saarland	1		−0.16	−0.17	0.14	.21	−0.45	0.10
Swabian	1		0.29	0.31	0.10	<.01	0.11	0.51
Turkish	1		−0.08	−0.09	0.26	.74	−0.60	0.43
Rater type	18	70%						
Heterogeneous convenience sample	8		−0.18	−0.21	0.08	<.01	−0.36	−0.06
Human resources managers	4		−0.65	−0.72	0.15	<.001	−1.01	−0.42
Students	6		0.16	0.17	0.09	.07	−0.01	0.35
Raters' relation to the study's standard variety	17	34%						
Native	7		−0.44	−0.50	0.13	<.001	−0.75	−0.24
Non-native	8		−0.03	−0.04	0.11	.68	−0.25	0.16
Mixed	2		0.08	0.08	0.21	.71	−0.34	0.49
Accent guise	18	0%						
Matched-guise	9		−0.16	−0.19	0.13	.13	−0.45	0.06
Verbal-guise	9		−0.14	−0.16	0.12	.20	−0.40	0.08
Match of applicants' and raters' language variety	17	18%						
No	10		−0.28	−0.32	0.11	<.01	−0.54	−0.10
Yes	2		0.24	0.26	0.23	.27	−0.20	0.72
Mixed	5		−0.08	−0.09	0.15	.52	−0.38	0.19

(Continues)

TABLE 3 (Continued)

Moderator	k	R^2	Hedges' g	δ	SE_δ	p	95% KI_δ LB	95% KI_δ UB
Any non-standard variety rater	13	0%						
Yes	10		−0.08	−0.10	0.09	.31	−0.28	0.09
Mixed	3		0.08	0.08	0.17	.62	−0.25	0.42

Note. All calculations are based on a random effects model; k = number of effect sizes included in the calculation; R^2 = variance explained in the heterogeneity of effect sizes by the model with significant moderators; Hedges' g = uncorrected estimate of Hedges' g; δ = estimate of Hedges' g corrected for criterion unreliability; SE_δ = standard error of the corrected estimate of Hedges' g; p = p-value of the corrected estimate of Hedges' g for the respective factor level; 95% KI_δ = 95% confidence interval of the corrected estimate of Hedges' g.
Abbreviations: AAVE = African American Vernacular English; LB = lower bound of the interval; UB = upper bound of the interval.

Moderator analyses of the relationship between language varieties and hirability

Finally, we investigated moderator effects for hirability (see Table 4) and found several significant moderators. First, the applicants' language variety type influenced hirability ratings, $Q_M(2) = 12.82$, $p < .01$, such that applicants speaking with a non-native language variety ($\delta = -0.74$, $p < .001$) were at a greater disadvantage in selection settings than applicants speaking with a native ethnic language variety ($\delta = -0.18$, $p = .13$). Second, not only the type of linguistic characteristic but also applicants' specific language variety itself played a role, $Q_M(21) = 61.38$, $p < .001$. Specifically, applicants were particularly disadvantaged if they spoke African American Vernacular English ($\delta = -0.72$, $p = .02$), or English with a Chinese ($\delta = -1.77$, $p < .001$), Hong Kong ($\delta = -1.28$, $p < .001$), Indian ($\delta = -1.16$, $p = .01$), Mexican ($\delta = -0.69$, $p < .001$), a strong German accent ($\delta = -0.87$, $p < .001$), or German with a Berlin ($\delta = -1.05$, $p = .02$) or Saxonian dialect ($\delta = -1.18$, $p < .001$). Third, the rater type also moderated the effect of non-native language, $Q_M(1) = 4.25$, $p = .04$, so that applicants evaluated by raters from heterogeneous convenience samples were more disadvantaged ($\delta = -0.65$, $p < .001$) than those evaluated by students ($\delta = -0.36$, $p < .001$) (however, it should be noted that none of the primary studies focusing on hirability as the dependent variable used human resources professionals as raters). Finally, concerning Research Question 3, we found that raters who spoke non-standard language themselves rated applicants speaking non-standard language lower ($\delta = -0.69$, $p < .001$) than did other raters ($\delta = -0.53$, $p = .02$ for raters who spoke standard language and $\delta = -0.26$, $p = .02$ for a mixed sample of raters), $Q_M(2) = 6.21$, $p = .04$. In contrast, the accent guise did not play a role for the hiring decisions, $Q_M(1) = 0.11$, $p = .74$. Likewise, neither the raters' relation to the study's standard variety, $Q_M(2) = 3.43$, $p = .18$, nor the match between language variety of applicants and raters, $Q_M(2) = 1.96$, $p = .38$, moderated the effects.

Publication bias analyses

The publication bias analyses were conducted separately for each dependent variable (competence, warmth, and hirability). Altogether, the evidence for publication bias was limited (for

TABLE 4 Analysis of the categorical moderators of the relationship between language variety and hirability.

Moderator	*k*	R^2	Hedges' *g*	δ	SE_δ	*p*	95% CI_δ LB	95% CI_δ UB
Applicants' language variety type	63	15%						
Native ethnic	21		−0.17	−0.18	0.12	.13	−0.41	0.05
Native regional	19		−0.57	−0.60	0.12	<.001	−0.84	−0.36
Non-native	23		−0.71	−0.74	0.11	<.001	−0.96	−0.52
Applicants' specific language variety	63	41%						
Standard language: Standard English								
AAVE	2		−0.70	−0.72	0.32	.02	−1.35	−0.09
Standard language: American English								
British English	1		0.14	0.15	0.45	.74	−0.74	1.03
Chinese	1		−1.69	−1.77	0.45	<.001	−2.66	0.88
Colombian	1		−0.11	−0.12	0.47	.80	−1.05	0.81
French	11		−0.09	−0.09	0.14	.49	−0.36	0.17
Hong Kong	3		−1.21	−1.28	0.28	<.001	−1.82	−0.73
Indian	1		−1.11	−1.16	0.45	.01	−2.05	−0.28
Japanese	8		−0.16	−0.16	0.16	.30	−0.47	0.15
Japanese and Russian	1		−0.55	−0.58	0.45	.20	−1.46	0.31
Mexican	6		−0.65	−0.69	0.18	<.001	−1.05	−0.33
German	2		−0.57	−0.59	0.33	.07	−1.23	0.05
Spanish	2		−0.46	−0.48	0.32	.14	−1.11	0.16
Strong German accent	3		−0.83	−0.87	0.26	<.01	−1.39	−0.35
Weak German accent	1		0.03	0.03	0.46	.95	−0.86	0.92
Standard language: British English								
Moderate Dutch accent	1		−0.32	−0.33	0.45	.46	−1.22	0.55
Slight Dutch accent	1		−0.33	−0.35	0.45	.44	−1.23	0.54
Standard language: Standard German								
Bavarian	7		−0.17	−0.18	0.17	.28	−0.51	0.15
Berlin	1		−1.00	−1.05	0.47	.02	−1.96	−0.13
Berlin and Saxon and Bavarian and Cologne	1		−0.49	−0.54	0.46	.24	−1.43	0.36
Saarland	1		−0.43	−0.45	0.46	.33	−1.36	0.45
Saxon	7		−1.13	−1.18	0.17	<.001	−1.51	−0.85
Swabian	1		0.00	0.00	0.45	1.00	−0.89	0.88
Rater type	63	5%						

(Continues)

TABLE 4 (Continued)

Moderator	k	R^2	Hedges' g	δ	SE_δ	p	95% CI_δ LB	95% CI_δ UB
Heterogeneous convenience sample	32		−0.62	−0.65	0.10	<.001	−0.85	−0.46
Students	31		−0.34	−0.36	0.10	<.001	−0.56	−0.16
Raters' relation to the study's standard variety	41	4%						
Native	8		−0.77	−0.81	0.22	<.001	−1.24	−0.38
Non-native	8		−0.52	−0.54	0.22	.01	−0.97	−0.11
Mixed	25		−0.33	−0.35	0.12	<.01	−0.59	−0.11
Accent guise	63	0%						
Matched-guise	8		−0.54	−0.57	0.21	.01	−0.98	−0.17
Verbal-guise	55		−0.48	−0.50	0.08	<.001	−0.66	−0.35
Match of applicants' and raters' language variety	35	0%						
No	11		−0.81	−0.85	0.18	<.001	−1.20	−0.50
Yes	2		−0.30	−0.31	0.42	.46	−1.13	0.51
Mixed	22		−0.58	−0.62	0.12	<.001	−0.86	−0.37
Any non-standard variety rater	49	8%						
No	6		−0.51	−0.53	0.23	.02	−0.98	−0.09
Yes	19		−0.65	−0.69	0.13	<.001	−0.94	−0.43
Mixed	24		−0.24	−0.26	0.11	.02	−0.48	−0.03

Note: All calculations are based on a random effects model; k = number of effect sizes included in the calculation; R^2 = variance explained in the heterogeneity of effect sizes by the model with significant moderators; Hedges' g = uncorrected estimate of Hedges' g; δ = estimate of Hedges' g corrected for criterion unreliability; SE_δ = standard error of the corrected estimate of Hedges' g; p = p-value of the corrected estimate of Hedges' g for the respective factor level; 95% CI_δ = 95% confidence interval of the corrected estimate of Hedges' g.
Abbreviations: AAVE = African American Vernacular English; LB = lower bound of the interval; UB = upper bound of the interval.

funnel plots, see Appendix F in the online supplemental materials). When the entire data set with all dependent variables was considered jointly, Egger's test for funnel plot asymmetry (Egger et al., 1997) did not detect publication bias, $t(107) = -1.61$, $p = .11$. Likewise, we found no bias for studies on competence, $t(26) = -0.71$, $p = .49$. However, for effect sizes on warmth, Egger's test indicated an asymmetry, $t(16) = -2.30$, $p = .03$, and the trim and fill method (Duval & Tweedie, 2000) suggested the imputation of two studies with a low inverse standard error and positive differences between non-standard and standard speaking applicants. The effect size estimate corrected for these potentially existing but unpublished studies was $\delta = -0.11$ ($p = .22$), which is slightly smaller and non-significant in comparison to the effect size from Table 1 ($\delta = -0.17$, $p = .04$). For hirability, Egger's test was also significant, $t(61) = -2.19$, $p = .03$, which suggests an asymmetry of the funnel plot. However, no imputation of studies was suggested.

DISCUSSION

Taken together, our results confirmed our Hypotheses 1 and 3 that applicants who are speaking non-standard language are perceived as less competent and less hirable than applicants who are speaking standard language, and the corresponding effect sizes were large for competence ($\delta = -0.70$) and moderate for hirability ($\delta = -0.51$). Furthermore, there was no evidence that publication bias affects these effect size estimates. In contrast to Hypothesis 2, which predicted a positive effect for warmth, we also found a negative but small effect of speaking non-standard language on warmth ($\delta = -0.17$). However, this effect turned non-significant when the estimation was corrected for publication bias. Taken together, our results are in line with previous meta-analytic findings of negative effects of non-standard language on perceptions of speakers (Fuertes et al., 2012) and extend these findings to personnel selection situations. They show that speaking a non-standard language puts applicants at a disadvantage in comparison to applicants who are speaking a standard language.

The negative effects of speaking non-standard language on perceptions of competence and hirability were true for the overwhelming majority of the different kinds of non-standard language and for perceptions of warmth in the case of native ethnic language varieties. However, results from the moderator analyses (even though these were often based on rather small numbers of primary studies) suggest that some language varieties are more detrimental than others. For instance, South-Asian and Indian accents (e.g., Timming, 2017) or Mexican accents (e.g., De la Zerda & Hopper, 1979) were perceived more negatively than others in the US and certain regional dialects such as Saxon or Berlin (e.g., Rakić et al., 2011) were perceived more negatively than others in Germany. One possible explanation for these results could be that there are particularly negative stereotypes about the respective language groups. Alternatively, it might also be that these accents or dialects appear more salient in comparison to the respective standard language and/or appear stronger than the accents and dialects in other studies. The finding that a strong German accent has a much more negative effect on competence and hirability ratings than a weak German accent when American English is the standard language (e.g., Roessel et al., 2019) would at least indirectly support such a possibility.

The moderator analyses showed that the meta-analytic results were also largely independent of the kind of raters from whom ratings were collected. Specifically, there was only limited evidence that human resources managers were less likely to be influenced by the effects of non-standard language than other groups of raters—and for warmth, they were more negatively affected. This finding parallels other research showing that experienced raters are also susceptible (even though possibly somewhat less) to the effects of biasing variables (e.g., Marlowe et al., 1996). Furthermore, with only two exceptions, we found no effects of whether the standard language was raters' mother tongue, whether raters spoke the same language variety as applicants, or whether raters spoke any non-standard variety themselves. The first exception in this regard was the finding that raters perceived applicants speaking with non-standard language as less warm than standard speakers when the standard variety of the respective study was their mother tongue but not if the raters had a different mother tongue. Given that we found significantly lower perceptions of warmth only for native ethnic language varieties, our evidence for a warmth-related bias primarily concerns situations when natives evaluate applicants with an ethnic background different from the majority or native population. The second exception was that raters speaking non-standard language themselves rated applicants with an accent or a dialect lower than other raters.

Concerning the experimental design, our moderator analyses also revealed that results from studies using the verbal-guise technique, in which standard and non-standard language were spoken by different individuals (e.g., Deprez-Sims & Morris, 2013), were largely similar in comparison to results from studies using the matched-guise technique, in which standard and non-standard language were spoken by the same individual (e.g., Hosoda et al., 2012). Even though the mean effect size for competence ratings from studies using the verbal-guise technique was descriptively larger than from studies using the matched-guise technique, this difference turned out to be non-significant. Furthermore, the meta-analytic effect sizes for warmth and hirability were rather similar. Thus, both experimental approaches lead to the same conclusions.

Limitations and lines for future research

There are several limitations that we want to discuss. These limitations concern the limited range of standard languages that were used in the available studies, the small number of studies for several of our moderator analyses, and the sole focus on interview settings in previous studies.

With regard to the first limitation, we only found primary studies in which either US/British English or German was used as the standard language. Thus, even though previous research investigated the effects of non-standard language in other language regions (see, for example, Yzerbyt et al., 2005, for the effects of a Belgian accent in comparison to standard French), this research did not consider these effects in selection contexts. This limits the generalizability of the present results.

Concerning the second limitation, the small number of primary studies for several of our moderator analyses affects both the analyses for specific language varieties and other variables. Regarding differences between different groups of raters, for example, we found no studies that investigated hirability as the dependent variable and used experienced recruiters as their participants. Given that experience or a relevant educational background can sometimes reduce the impact of biases (e.g., Marlowe et al., 1996) and contribute to the validity of ratings (e.g., Gaugler et al., 1987), it would be good to have primary studies with non-student and non-convenience samples.

Another limitation concerned the sole focus of the different primary studies on interviews. Even though we had initially also searched for studies with assessment centers, we only found studies with interviews. Furthermore, all these interviews were structured only moderately at best.[2] Specifically, most used traditional interview questions and did not provide training or rating anchors for the study participants who had to rate the applicants in the interviews. Thus, the effects of non-standard language may have been smaller if additional aspects of interview structure and more job-related interview questions had been used (cf. Campion et al., 1997), given that the effects of potentially biasing factors such as applicants' impression management and attractiveness (e.g., Barrick et al., 2009; Bill et al., 2023), obesity (Kutcher & Bragger, 2004), ethnicity (Huffcutt & Roth, 1998), or sex (Huffcutt et al., 2001) are considerably smaller in highly structured interviews than in less structured interviews. Accordingly, future research with structured interviews and with other selection procedures such as assessment centers is needed.

Another restriction concerning the primary studies was that none of them were from high-stakes field settings. Thus, the generalizability to actual field settings also needs to be evaluated

in future research. It is also unclear how non-standard language affects the criterion-related validity of interviews. If only some applicants speak non-standard language, this could increase error variance in interview ratings, which could eventually impair interview validity. However, research is necessary to evaluate the actual impact of non-standard language on interview validity.

Practical implications

This meta-analysis highlights the influence of language varieties on personnel selection, emphasizing the need to confront these biases head-on. Applicants proficient in standard language may find it beneficial to speak this standard language in selection contexts. Likewise, job seekers who do not speak standard language may choose pronunciation training in standard language (e.g., O'Brien, 2021) to increase the likelihood of receiving a job offer.

However, the onus is on organizations to adopt inclusive practices that ensure fair evaluation of all candidates. As already noted above, one approach that seems especially viable in this regard is the use of structured interviews. Specifically, if one uses a set of predefined, job-relevant questions for all applicants, then this already reduces potential biases (e.g., Barrick et al., 2009). Furthermore, if applicants' answers are evaluated according to a common frame of reference, then this further reduces biasing factors (e.g., Kutcher & Bragger, 2004). To do so, either providing descriptive rating anchors for each question or conducting a so-called frame-of-reference training for raters seem to be strategies that are comparable with regard to their effects on rating accuracy (Melchers et al., 2011). Furthermore, given that structured interviews belong to the most valid selection tools (Sackett et al., 2022), using these measures does not only promise to limit the biasing impact of non-standard language but also fosters the overall quality of the selection process.

ACKNOWLEDGEMENTS

Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

This meta-analysis was conducted in accordance with ethical standards for research synthesis. As this study involved the aggregation and analysis of data from previously published studies, no direct data collection from participants was undertaken, and thus, approval from an ethics committee was not applicable. The compiled dataset and further materials used for this meta-analysis are available at https://osf.io/zrv73/.

ETHICS STATEMENT

This meta-analysis was conducted in accordance with ethical standards for research synthesis. As this study involved the aggregation and analysis of data from previously published studies, no direct data collection from participants was undertaken, and thus, approval from an ethics committee was not applicable.

ORCID

Niklas Schulte https://orcid.org/0000-0001-8915-7503
Johannes M. Basch https://orcid.org/0000-0003-4847-2181
Klaus G. Melchers https://orcid.org/0000-0003-4211-6450

ENDNOTES

[1] We also conducted analyses for the continuous moderator variables percentage of female applicants, raters' mean age, percentage of raters being migrants, and the percentage of raters with experience in applicant selection (see Appendix E in the online supplemental materials). None of them showed significant influence on effect sizes except raters' mean age and the percentage of experienced raters, which were negatively related to the effects sizes for warmth.

[2] To avoid potential misunderstandings, we would like to point out that the stimulus materials in the experimental studies were standardized. However, interview structure does not refer to standardization of these materials but rather to the content of the interview (i.e., what is asked in the different questions) and the way how answers are evaluated (e.g., according to a common evaluative standard, see Campion et al., 1997, for an extensive review of the meaning of interview structure).

REFERENCES

References marked with an asterisk indicate studies included in the meta-analysis.

Abbasi, Z., Billsberry, J., & Todres, M. (2023). Empirical studies of the "similarity leads to attraction" hypothesis in workplace interactions: A systematic review. *Management Review Quarterly*, advance online publication. https://doi.org/10.1007/s11301-022-00313-5

Agerström, J., Björklund, F., Carlsson, R., & Rooth, D.-O. (2012). Warm and competent Hassan = cold and incompetent Eric: A harsh equation of real-life hiring discrimination. *Basic and Applied Social Psychology*, *34*(4), 359–366. https://doi.org/10.1080/01973533.2012.693438

Amaral, A. A., Powell, D. M., & Ho, J. L. (2019). Why does impression management positively influence interview ratings? The mediating role of competence and warmth. *International Journal of Selection and Assessment*, *27*(4), 315–327. https://doi.org/10.1111/ijsa.12260

Ammon, U. (2004). Standard Variety/Standardvarietät. In U. Ammon, N. Dittmar, K. J. Mattheier, & P. Trudgill (Eds.), *Sociolinguistics/Soziolinguistik: An international handbook of the science of language and society/Ein internationales Handbuch zur Wissenschaft von Sprache und Gesellschaft*. De Gruyter Mouton. https://doi.org/10.1515/9783110141894.1.2.273

*Bald, J., & Kanning, U. P. (2019). Urteilsverzerrungen in der personalauswahl–Schneiden bewerberinnen, die mit akzent sprechen, im interview schlechter ab und welche rolle spielt die erfahrung der entscheidungsträger? [Bias in personnel selection—Do applicants that speak with an accent lower interview ratings and which role does interviewer experience play?]. *Journal of Business and Media Psychology*, *10*(1), 49–56.

Barrick, M. R., Shaffer, J. A., & DeGrassi, S. W. (2009). What you see may not be what you get: Relationships among self-presentation tactics and ratings of interview and job performance. *Journal of Applied Psychology*, *94*(6), 1394–1411. https://doi.org/10.1037/a0016532

Bartkoski, T., Lynch, E., Witt, C., & Rudolph, C. (2018). A meta-analysis of hiring discrimination against Muslims and Arabs. *Personnel Assessment and Decisions*, *4*(2), 1–16. https://doi.org/10.25035/pad.2018.02.001

Batinovic, L., Howe, M., Sinclair, S., & Carlsson, R. (2023). Ageism in hiring: A systematic review and meta-analysis of age discrimination. *Collabra: Psychology*, *9*(1). https://doi.org/10.1525/collabra.82194

Bill, B., Melchers, K. G., Steuer, J., & Eisele, E. (2023). Are traditional interviews more prone to effects of impression management than structured interviews? *Applied Psychology. An International Review*. Advance online publication. https://doi.org/10.1111/apps.12514

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Wiley. https://doi.org/10.1002/9780470743386

Bradac, J. J., & Wisegarver, R. (1984). Ascribed status, lexical diversity, and accent: Determinants of perceived status, solidarity, and control of speech style. *Journal of Language and Social Psychology*, *3*(4), 239–255. https://doi.org/10.1177/0261927X8400300401

Byrne, D. E. (1971). *The attraction paradigm*. Academic Press.

Campion, M. A., Palmer, D. K., & Campion, J. E. (1997). A review of structure in the selection interview. *Personnel Psychology*, *50*(3), 655–702. https://doi.org/10.1111/j.1744-6570.1997.tb00709.x

*Cargile, A. C. (1997). Attitudes toward Chinese-accented speech: An investigation in two contexts. *Journal of Language and Social Psychology*, *16*(4), 434–443. https://doi.org/10.1177/0261927X97016400

Chen, C. C., & Lin, M. M. (2014). The effect of applicant impression management tactics on hiring recommendations: Cognitive and affective processes. *Applied Psychology. An International Review*, *63*(4), 698–724. https://doi.org/10.1111/apps.12013

Cheyne, W. M. (1970). Stereotyped reactions to speakers with Scottish and English regional accents. *The British Journal of Social and Clinical Psychology*, *9*(1), 77–79. https://doi.org/10.1111/j.2044-8260.1970.tb00642.x

*Chui, T. S., & Lee, Y. T. (2015). *The effect of English accents on hiring recommendations. Is "ABC" accent more preferred in Hong Kong?* [Unpublished undergraduate thesis, Hong Kong Baptist University].

*Cocchiara, F. K. (2007). *Sociolinguistic cues, perceived race and employment selection outcomes: An exploration of the aversive racism framework* [Unpublished doctoral thesis, University of Texas at Arlington].

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Cuddy, A. J., Glick, P., & Beninger, A. (2012). The dynamics of warmth and competence judgments, and their outcomes in organizations. *Research in Organizational Behavior*, *31*, 73–98. https://doi.org/10.1016/j.riob.2011.10.004

De la Zerda, N., & Hopper, R. (1979). Employment interviewers' reactions to Mexican American speech. *Communication Monographs*, *46*(2), 126–134. https://doi.org/10.1080/03637757909375998

*Deprez-Sims, A. S., & 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*(4), 355–367. https://doi.org/10.1111/ijsa.12045

*Duroska, S., Hutmacher, D., Maes, E., Nesser, M., & Ramsperger, S. (2014). Dialekt als karrierebremse? Einfluss des sprachstils auf einstellungschancen im beruf [Dialect as a career stopper? The influence of speech style on the chance to get hired]. *Luxemburger Experimentalpraktikum Journal*, *8*, 15–24.

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455–463. https://doi.org/10.1111/j.0006-341X.2000.00455.x

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, *315*, 629–634. https://doi.org/10.1136/bmj.315.7109.629

Fetscherin, M., Tantleff-Dunn, S., & Klumb, A. (2020). Effects of facial features and styling elements on perceptions of competence, warmth, and hireability of male professionals. *The Journal of Social Psychology*, *160*(3), 332–345. https://doi.org/10.1080/00224545.2019.1671300

Fiske, S. T. (2004). Intent and ordinary bias: Unintended thought and social motivation create casual prejudice. *Social Justice Research*, *17*(2), 117–127. https://doi.org/10.1023/B:SORE.0000027405.94966.23

Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*(6), 878–902. https://doi.org/10.1037//0022-3514.82.6.878

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. *Advances in Experimental Social Psychology*, *23*, 1–74. https://doi.org/10.1016/S0065-2601(08)60317-2

Fuertes, J. N., Gottdiener, W. H., Martin, H., Gilbert, T. C., & Giles, H. (2012). A meta-analysis of the effects of speakers' accents on interpersonal evaluations. *European Journal of Social Psychology*, *42*(1), 120–133. https://doi.org/10.1002/ejsp.862

Gaugler, B. B., Rosenthal, D. B., Thornton, G. C. III, & Bentson, C. (1987). Meta-analysis of assessment center validity. *Journal of Applied Psychology*, *72*(3), 493–511. https://doi.org/10.1037/0021-9010.72.3.493

Giles, H. (1970). Evaluative reactions to accents. *Educational Review*, *22*(3), 211–227. https://doi.org/10.1080/0013191700220301

Giles, H. (1971). Patterns of evaluation to R.P., South Welsh and Somerset accented speech. *The British Journal of Social and Clinical Psychology*, *10*(3), 280–281. https://doi.org/10.1111/j.2044-8260.1971.tb00748.x

Giles, H., & Billings, A. C. (2004). Assessing language attitudes: Speaker evaluation studies. In A. Davies & C. Elder (Eds.), *The handbook of applied linguistics* (pp. 187–209). Blackwell. https://doi.org/10.1002/9780470757000.ch7

Giles, H., & Sassoon, C. (1983). The effect of speaker's accent, social class background and message style on British listeners' social judgements. *Language & Communication*, *3*(3), 305–313. https://doi.org/10.1016/0271-5309(83)90006-X

Giles, H., Wilson, P., & Conway, A. (1981). Accent and lexical diversity as determinants of impression formation and perceived employment suitability. *Language Sciences*, *3*(1), 91–103. https://doi.org/10.1016/S0388-0001(81)80015-0

Goldberg, C. B. (2005). Relational demography and similarity–attraction in interview assessments and subsequent offer decisions: Are we missing something? *Group & Organization Management*, *30*(6), 597–624. https://doi.org/10.1177/1059601104267661

*Hansen, K., Rakić, T., & Steffens, M. C. (2017). Competent and warm? How mismatching appearance and accent influence first impressions. *Experimental Psychology*, *64*(1), 27–36. https://doi.org/10.1027/1618-3169/a000348

Hernández, M., Ventura-Campos, N., Costa, A., Miró-Padilla, A., & Ávila, C. (2019). Brain networks involved in accented speech processing. *Brain and Language*, *194*, 12–22. https://doi.org/10.1016/j.bandl.2019.03.003

Horverak, J., Sandal, G., Bye, H., & Pallesen, S. (2013). Managers' selection preferences: The role of prejudice and multicultural personality traits in the assessment of native and immigrant job candidates. *European Review of Applied Psychology*, *63*(5), 267–275. https://doi.org/10.1016/j.erap.2013.07.003

*Hosoda, M., Nguyen, L. T., & Stone-Romero, E. F. (2012). The effect of Hispanic accents on employment decisions. *Journal of Managerial Psychology*, *27*(4), 347–364. https://doi.org/10.1108/02683941211220162

*Hosoda, M., & Stone-Romero, E. (2010). The effects of foreign accents on employment-related decisions. *Journal of Managerial Psychology*, *25*, 113–132. https://doi.org/10.1108/02683941011019339

Huang, L., Frideger, M., & Pearce, J. L. (2013). Political skill: Explaining the effects of nonnative accent on managerial hiring and entrepreneurial investment decisions. *Journal of Applied Psychology*, *98*(6), 1005–1017. https://doi.org/10.1037/a0034125

Huffcutt, A. I., Conway, J. M., Roth, P. L., & Stone, N. J. (2001). Identification and meta-analytic assessment of psychological constructs measured in employment interviews. *Journal of Applied Psychology*, *86*(5), 897–913. https://doi.org/10.1037/0021-9010.86.5.897

Huffcutt, A. I., & Roth, P. L. (1998). Racial group differences in employment interview evaluations. *Journal of Applied Psychology*, *83*(2), 179–189. https://doi.org/10.1037/0021-9010.83.2.179

*Korsten, M. (2020). *The effect of Dutch-accented English on the evaluations of job candidates by nonnative listeners in a hiring process* [Unpublished undergraduate thesis, Radboud University Nijmegen].

Kutcher, E. J., & Bragger, J. D. (2004). Selection interviews of overweight job applicants: Can structure reduce the bias? *Journal of Applied Social Psychology*, *34*(10), 1993–2022. https://doi.org/10.1111/j.1559-1816.2004.tb02688.x

Lambert, W. E., Hodgson, R. C., Gardner, R. C., & Fillenbaum, S. (1960). Evaluational reactions to spoken languages. *Journal of Abnormal and Social Psychology*, *60*(1), 44–51. https://doi.org/10.1037/h0044430

Lievens, F. (2001). Assessor training strategies and their effects on accuracy, interrater reliability, and discriminant validity. *Journal of Applied Psychology*, *86*(2), 255–264. https://doi.org/10.1037/0021-9010.86.2.255

Lim, S. L., Bentley, P. J., Peterson, R. S., Hu, X., & Prouty McLaren, J. (2023). Kill chaos with kindness: Agreeableness improves team performance under uncertainty. *Collective Intelligence*, *2*(1), 263391372311585. https://doi.org/10.1177/26339137231158584

Lippens, L., Vermeiren, S., & Baert, S. (2023). The state of hiring discrimination: A meta-analysis of (almost) all recent correspondence experiments. *European Economic Review*, *151*, 104315. https://doi.org/10.1016/j.euroecorev.2022.104315

Mai, R., & Hoffmann, S. (2010). Die wirkung von akzent und dialekt in der internen und externen kommunikation: Stand der betriebswirtschaftlich orientierten forschung und forschungsdirektiven [The

effect of accent and dialect on internal and external communication: State of the art in economically oriented research and lines for future research]. *Journal für Betriebswirtschaft*, *60*(4), 241–268. https://doi.org/10.1007/s11301-010-0065-7

Marlowe, C. M., Schneider, S. L., & Nelson, C. E. (1996). Gender and attractiveness biases in hiring decisions: Are more experienced managers less biased? *Journal of Applied Psychology*, *81*(1), 11–21. https://doi.org/10.1037/0021-9010.81.1.11

Martins, L. L., Gilson, L. L., & Maynard, M. T. (2004). Virtual teams: What do we know and where do we go from here? *Journal of Management*, *30*(6), 805–835. https://doi.org/10.1016/j.jm.2004.05.002

McAuliffe, M., & Triandafyllidiou, A. (Eds.) (2021). *World migration report 2022*. International Organization for Migration (IOM). https://publications.iom.int/books/world-migration-report-2022

Melchers, K. G., Lienhardt, N., von Aarburg, M., & Kleinmann, M. (2011). Is more structure really better? A comparison of frame-of-reference training and descriptively anchored rating scales to improve interviewers' rating quality. *Personnel Psychology*, *64*(1), 53–87. https://doi.org/10.1111/j.1744-6570.2010.01202.x

*Michelsen, K. (2019). *Dialect and employability: Human resource managers' perceptions of African American English* [Unpublished undergraduate thesis, Brigham Young University].

*Nijs, N. D. (2019). *The effects of non-native accents on hirability: A comparison of German-accented English, Spanish-accented English and American English in job interviews* [Unpublished undergraduate thesis, Radboud University Nijmegen].

O'Brien, M. G. (2021). Ease and difficulty in L2 pronunciation teaching: A mini-review. *Frontiers in Communication*, *5*, 626985. https://doi.org/10.3389/fcomm.2020.626985

R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.R-project.org/.

*Rakić, T., Steffens, M. C., & Mummendey, A. (2011). When it matters how you pronounce it: The influence of regional accents on job interview outcome. *British Journal of Psychology*, *102*(4), 868–883. https://doi.org/10.1111/j.2044-8295.2011.02051.x

*Roessel, J., Schoel, C., Zimmermann, R., & Stahlberg, D. (2019). Shedding new light on the evaluation of accented speakers: Basic mechanisms behind nonnative listeners' evaluations of nonnative accented job candidates. *Journal of Language and Social Psychology*, *38*(1), 3–32. https://doi.org/10.1177/0261927X17747904

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis* (Vol. 621, pp. 231–244). Russell Sage Foundation.

Sackett, P. R., Zhang, C., Berry, C. M., & Lievens, F. (2022). Revisiting meta-analytic estimates of validity in personnel selection: Addressing systematic overcorrection for restriction of range. *Journal of Applied Psychology*, *107*(11), 2040–2068. https://doi.org/10.1037/apl0000994

Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist*, *32*(9), 752–760. https://doi.org/10.1037/0003-066X.32.9.752

*Sterk, R. (2020). *Du bist was du sprichst - Der Einfluss des schwäbischen Dialekts auf die Beurteilung in Vorstellungsgesprächen* [Unveröffentlichte Bachelorarbeit, Universität Ulm].

Tajfel, H. (1970). Experiments in intergroup discrimination. *Scientific American*, *223*(5), 96–103. https://doi.org/10.1038/scientificamerican1170-96

Tajfel, H. (1974). Social identity and intergroup behaviour. *Social Science Information*, *13*(2), 65–93. https://doi.org/10.1177/053901847401300204

Tajfel, H., & Wilkes, A. L. (1963). Classification and quantitative judgement. *British Journal of Psychology*, *54*(2), 101–114. https://doi.org/10.1111/j.2044-8295.1963.tb00865.x

Tasa, K., Sears, G. J., & Schat, A. C. (2011). Personality and teamwork behavior in context: The cross-level moderating role of collective efficacy. *Journal of Organizational Behavior*, *32*(1), 65–85. https://doi.org/10.1002/job.680

*Timming, A. R. (2017). The effect of foreign accent on employability: A study of the aural dimensions of aesthetic labour in customer-facing and non-customer-facing jobs. *Work, Employment and Society*, *31*(3), 409–428. https://doi.org/10.1177/0950017016630260

Varghese, L., Lindeman, M. I. H., & Finkelstein, L. (2018). Dodging the double bind: The role of warmth and competence on the relationship between interview communication styles and perceptions of women's hirability. *European Journal of Work and Organizational Psychology*, *27*(4), 418–429. https://doi.org/10.1080/1359432X.2018.1463989

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

*Visscher, C. R. (2019). *Hiring success: How a non-native accent affects the evaluation of a non-native speaker in job interviews* [Unpublished undergraduate thesis, Radboud University Nijmegen].

Yule, G. (2006). *The study of language* (3rd ed.). Cambridge University Press.

Yzerbyt, V., Provost, V., & Corneille, O. (2005). Not competent but warm... really? Compensatory stereotypes in the French-speaking world. *Group Processes & Intergroup Relations*, *8*(3), 291–308. https://doi.org/10.1177/1368430205053944

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, *9*(2), 1–27. https://doi.org/10.1037/h0025848

How to cite this article: Schulte, N., Basch, J. M., Hay, H.-S., & Melchers, K. G. (2024). Do ethnic, migration-based, and regional language varieties put applicants at a disadvantage? A meta-analysis of biases in personnel selection. *Applied Psychology*, *73*(4), 1866–1892. https://doi.org/10.1111/apps.12528