The influence of ethnicity and language variation on undergraduates' evaluations of Dutch-speaking instructors in Belgium: a contextualized speaker evaluation experiment

Highlights

- Standard language ideological beliefs are weak in a contextualised experiment.
- Standard-speaking Maghrebi instructors are not downgraded compared to Flemings.
- The standard-speaking instructor with hijab excelled in professionalism.

Abstract

This study reports on a contextualised speaker evaluation experiment exploring the effects of language variation (standard vs. colloquial varieties), ethnic identity (Flemish vs. Maghrebi) and wearing a headscarf on students' evaluation of a female university instructor in Flanders. 314 participants evaluated a single lecture of the instructor on professionalism, comprehensibility, authority, standardness, social attractiveness and physical attractiveness. The results suggest that when measured indirectly, colloquial Dutch is no less acceptable than Standard Dutch in higher education. No general downgrading of the Maghrebi identity was observed. When wearing a headscarf, the standard-speaking Maghrebi instructor was upgraded for professionalism.

Keywords

Speaker evaluation experiment, language and discrimination, ethnic stereotyping, education, language ideology

1. Introduction

In recent years, language attitude research has fruitfully applied experimental methods to the study of language ideology and language ideological change. Such methods often deploy some form of speaker evaluation experiment, such as the matched-guise technique (Lambert et al., 1960), adopted from social psychology, in which participants rate one bilingual or bidialectal speaker in different 'guises' on several personality traits. Speaker evaluation experiments deploying this technique, or variants of it, have mostly demonstrated robust and enduring associations of standard languages with power- and status-related traits, such as competence, intelligence, ambition and professional success, and of vernacular varieties with solidarity, social attractiveness and likeability (Garrett, 2010; Grondelaers, 2013). These findings have been taken as evidence for the uptake of a standard language ideology, at the core of which is a hierarchisation of standard languages as superior and of vernacular varieties as non-prestigious and homely (Agha, 2003; Garrett, 2010; Jaspers and Van Hoof, 2013). Recently, speaker evaluation experiments have moreover managed to elicit 'dynamism' perceptions for some vernacular varieties, suggesting that these might be the driving force behind a language ideological change in which these vernacular varieties are acquiring a new type of prestige, viz. a cool and credibility deriving from their usage in the media (Grondelaers and Speelman, 2013; Kristiansen, 2001; 2009).

By explicitly or implicitly equating language attitudes to the social meanings or indexicality of speech styles (Grondelaers et al., 2016; Rosseel et al., 2019a,b; Soukup, 2013), and by treating them as indicative for the vitality, or on-going change, of language ideologies, quantitative language attitude studies have gone through a conceptual rapprochement with interpretive and qualitative strands of sociolinguistics and with linguistic anthropology, the field where the notions of language ideology and indexicality were originally developed (Kroskrity, 2010; Silverstein, 1979; 2003; Woolard and Schieffelin, 1994). It has also been noted, however, that experimental methodologies sit uncomfortably with the contextualized nature of social meaning that is foregrounded in these fields (Potter & Wetherell, 1987; Soukup, 2013; Rosseel et al., 2019a). Indeed, central there is the observation that the meanings of linguistic variables that mark standard

and vernacular varieties 'are not precise or fixed but rather constitute a field of potential meanings – an indexical field, or constellation of ideologically related meanings, any one of which can be activated in the situated use of the variable' (Eckert, 2008). In a similar vein, Coupland has warned sociolinguists against 'oversimplif[ying their] account[s] of the ideological loading of linguistic varieties', on the grounds that 'contextual factors impinge crucially on which social meanings are attributed to linguistic varieties' (2010, p. 134). Moreover, he has pointed out that when studied *in situ*, the indexicality of varieties works in tandem with 'the semiotics of movement, body shape and stature, physical and physiognomic beauty, clothing, and so on' (2010, p. 144).

To be sure, the multimodality of meaning generation has not gone unnoticed in experimental studies, as also research in this tradition has pointed out that the perception of a speaker's social identity, indexed by 'speakers' skin colour, dress, and so forth' (Dragojevic, et al., 2018, p. 17), may cue the perception of that speaker's language style. In such a process of 'reverse linguistic stereotyping', as Kang and Rubin (2009) have called it, it is not the stereotypes ideologically associated with the speech style that influence the perception of the speaker, but the other way around: attributions of a speaker's group membership and the stereotypes attached to that group influence the assessment and comprehension of their speech style. Thus, in a series of matched-guise experiments in which American students evaluated instructors, Rubin and colleagues found that when listeners saw a speaker belonging to an ethnic minority and were led to believe they were listening to a non-native speaker of English, they reported hearing a non-native accent when in fact there was none, their listening comprehension declined significantly and the instructor's teaching skills were evaluated as poorer (Kang and Rubin, 2009; Rubin, 1992; Rubin et al., 1999; Rubin and Smith, 1990). Several other studies have confirmed that seeing a foreign-looking face can make a speaker sound more foreign-accented (Yi et al., 2013; Yi et al., 2014; Zheng and Samuel, 2017).

Findings such as these demonstrate the interplay between language attitudes and ethnic bias and thus confirm that it is worth heeding Rosseel et al.'s call to 'try and include context features in [...] experiments on language attitudes [...] in order to obtain results that are more ecologically valid and tell us more about how language attitudes function in society' (2019a, p. 454), and, we might add, to get into view the impact of language ideology in specific contexts. In line with this call, the purpose of this paper is to further explore the potential of a contextualized speaker evaluation design, in order to investigate whether the attitudes typical of a standard language ideology can also be elicited in a situated context (instead of *in abstracto*), and whether these attitudes may also be affected (mitigated, heightened, or overruled) by speaker characteristics.

We developed a contextualised speaker evaluation experiment situating the speaker in an educational setting, where we expect the workings of a standard language ideology to be particularly salient. Indeed, in Flanders, the officially Dutch-speaking part of Belgium, where we conducted our experiment, teachers are widely seen as 'guardians of the standard language' (Van De Velde and Houtermans, 1999), i.e. exemplary speakers who are supposed to uphold and spread the standard language norm (Delarue, 2013; Grondelaers and Van Hout, 2012, p. 48). However, no attitudinal data are as yet available demonstrating that students do indeed expect instructors to live up to that ideal and downgrade nonstandard usage. Similar to Rubin (1992), we opted for a between-subjects-design, in which several groups of undergraduate students evaluated a single instructor, presented to them in different guises which consisted of both an auditory and a visual stimulus. Each group heard one version of a fragment of a lecture which was combined with names and pictures of the same instructor, representing different ethnic minority vs. majority identities. In addition, we factored in standardness of speech as a variable, in line with the 'classic' speaker evaluation experiments that probe for differential assessments of standard and vernacular styles. Whereas most speaker evaluation studies deploy male speakers, in this study we opted for a female speaker, which allowed us to incorporate an extra marker of a minority identity, notably one that indexes religious affiliation, viz., a headscarf.

As a result, this study was guided by the following research questions:

RQ1: what attitudes do we find vis à vis standard and vernacular varieties in an educational setting, a context which is traditionally taken to be standard language territory?

RQ2: how does ethnic identity impact on speaker evaluation in this context?

RQ3: how does wearing a headscarf impact on speaker evaluation in this context?

Before further elaborating on the design of our study (section 3), we will discuss our research context and previous studies on language attitudes and ethnic discrimination conducted in that context in section 2. The results of our study will be discussed in section 4; in the final sections we elaborate on the implications of our findings.

2. Language attitudes and ethnic discrimination in Flanders

It is often claimed that Flanders is characterized by a particularly strong standard language ideology, as a result of a heavy investment, particularly in the post-WWII decades, of education, the media and numerous grassroots organizations in the spread and promotion of Standard Dutch (for further details, see Jaspers and Van Hoof, 2013). As a result, Belgian Standard Dutch today unproblematically functions as the norm for written communication, exhibiting little difference with Standard Netherlandic Dutch. In many, mostly spoken, domains, however, Flemish language use deviates markedly from the standard. While dialect use is on the wane in all Flemish provinces, including the westernmost peripheral areas in which it has for a long time been most vital (Vandekerckhove, 2009), hybrid colloquial styles are on the rise. They are captured by the umbrella term *tussentaal* (literally 'in-between-language'), referring to the 'intermediate' stratificational position of its structural ingredients, in between the Flemish dialects and Belgian Standard Dutch. As the synonym 'Colloquial (Belgian) Dutch' that some linguists have adopted for *tussentaal* indicates, the variety mostly surfaces in conversations in informal (private and unofficial) contexts, although its features can also increasingly be heard on radio and television, notably in entertainment genres, while serious genres such as the news and current affairs shows remain the exclusive preserve of Belgian Standard Dutch (Van Hoof, 2018). Despite the unmistakable vitality of Colloquial Belgian Dutch, many opinion makers and intellectuals still put the premium on Standard Dutch as the only acceptable variety that ought to be used and promoted in the media and in education, and the only variety that will ensure equal opportunities, upward social mobility, and societal inclusion, notably for newcomers and members of ethnic minorities (Jaspers, 2012; 2013).

In order to gauge to what extent Flemings have internalised this pro-standard metadiscourse, sociolinguists have since the 1970s conducted attitudinal studies, which initially took the form of questionnaires (Deprez, 1981; Geerts, et al., 1980; Meeus, 1974; Vandekerckhove, 2000) and later made use of speaker evaluation techniques developed in the field of social psychology, notably the matched guise technique (or related variants – Impe and Speelman, 2007; Ghyselen, 2009; Vandekerckhove and Cuvelier, 2007; Grondelaers and Speelman, 2013). In this technique, one speaker is recorded in different 'guises', i.e. using different speech styles. These are then evaluated by respondents, who are asked to rate 'each speaker' on a variety of personality traits and are thus left unaware that they are in fact rating the same speaker. Usually no language-related questions are asked, so that the exact object of the study remains unclear. More recently, researchers have adopted alternative experimental methods, such as the Implicit Association Test and the Relational Response Task, which are both reaction time-based measures developed in social psychology (Rosseel et al., 2019a,b), or free response tasks analysed with big-data-based techniques (Grondelaers et al., 2020). In general, the abovementioned studies indicate that the status of Standard Dutch has remained largely uncontested. Colloquial Belgian Dutch is upgraded on solidarity and dynamism (i.e. trendiness and assertiveness), but generally downgraded on traditional prestige features, power and competence, which seem to remain the exclusive territory of Standard Dutch. The abovementioned studies thus seem to confirm Geeraerts and Van de Velde's (2013, p. 539) assertion that 'Colloquial Belgian Dutch [...] is not a new norm for Belgian Dutch per se, but it occupies a specific position as a language of what could be identified

as “public interaction in a non-professional, entertainment-oriented sphere”. Belgian Standard Dutch accordingly involves public interaction in a professional, expertise-oriented sphere, while the dialects are associated with personal interaction in the private sphere.’ Some researchers have interpreted this division of labour as a sign of the enduring vitality of the standard language ideology (Jaspers and Van Hoof, 2013; Rosseel et al., 2019a). Others have interpreted the public media licensing of Colloquial Dutch and the association of the variety with dynamism as a sign of changes in the traditional standard language ideology (Grondelaers and Speelman, 2013; Grondelaers et al., 2016).

These contrasting interpretations set aside, most of the speaker evaluation experiments mentioned above had speakers assessed *in abstracto* and thus laid bare decontextualized attitudes towards Belgian Standard Dutch and Colloquial Dutch (notable exceptions being Ghyselen, 2009, whose speakers were presented to the participants as teachers, and Vandekerckhove and Cuvelier, 2007, whose participants assessed audio-visually recorded scenes in shopping contexts). Flemish experimental research into language attitudes has thus as yet had little eye for the possible impact of contextual and speaker characteristics on speaker assessment (cf. Rosseel et al., 2019a). Almost all speaker evaluation experiments have deployed male speakers, neglecting the potential impact of gender on evaluation outcomes. The almost exclusive focus on ethnic majority speakers and the neglect of the potential impact of ethnicity in Flemish studies on language attitudes is somewhat remarkable too, given that over the past few decades processes of globalisation and international migration have greatly increased ethnic, cultural and linguistic heterogeneity in Flanders (but see Marzo, 2016 as a notable exception). Today 20.5% of residents in Flanders has a foreign background (i.e. has or had a foreign nationality or has at least one parent with a foreign nationality). Moroccans and Turks constitute the two largest minority groups of non-EU descent in Flanders (Noppe et al., 2018; Myria, 2017). The impact of ethnic bias in particular on speaker evaluations has thus become all the more relevant to assess. The experimental study conducted by Grondelaers and Van Gent (2019) in the Netherlands revealed that accent strength impacted differently on speaker evaluations, depending on whether the speaker belonged to the ethnic majority or had a migrant background. The male speakers in their experiment spoke with ‘indigenous’ regional accents, more specifically from the Dutch provinces of Limburg and Brabant, or with Moroccan accents to varying degrees, and were identifiable as ethnic minority or majority members on the basis of their names. Whereas the reduction of a regional accent boosted the scores of Limburgish speakers with a Dutch name on the superiority dimension, speakers with Arabic names were consistently downgraded on this dimension compared to the speakers of indigenous varieties with Dutch names. Although speakers with Arabic names and mild and strong Moroccan accents scored lowest on superiority, speakers with an Arabic name and no discernible Moroccan accent were still downgraded on the superiority dimension, from which Grondelaers and Van Gent inferred that ‘unaccentedness does not compensate for [the] stigma’ of an Arabic name (2019, p. 8).

It is far from unthinkable that perceptions of ethnicity influence the evaluation of ethnic minority speakers negatively in Flanders as well, since scholars in the fields of labour economics, sociology and personnel psychology have provided ample evidence that ethnic discrimination is rife, inter alia on the labour market and the rental housing market. In so-called correspondence tests, in which fictitious job applications or requests for a viewing appointment are sent out in response to real vacancies for jobs or rental houses, unequal treatment on the basis of an ethnic minority background has repeatedly been laid bare (Baert, 2018; Baert et al., 2015; Baert and Vujić, 2016; Baert et al., 2017; Verhaeghe and Platform Praktijktesten nu, 2018; Heylen and Van den Broeck, 2016). Language has received little attention as a possible discrimination ground in these studies. However, in their telephone survey conducted on the rental housing market in the cities of Antwerp and Ghent, Van der Bracht et al. (2015) found that migrants who are proficient in Dutch and have no noticeable foreign accent are discriminated against as often as migrants with lower proficiency in Dutch.

Discrimination on ethnic grounds is moreover fuelled by a climate in which there is a general distrust of Muslims, who constitute a large part of ethnic minorities in Flanders, in particular. Typical for this climate are the heated debates that have been taking place for several years now on the right to express one’s

religious identity at school or at work, notably through wearing a hijab. The question whether refusing access to certain domains or positions on the basis of wearing a headscarf constitutes discrimination is hotly contested, which is probably one of the reasons why there is still little empirical evidence for discrimination against Muslim women in Flanders. At the same time, the lack of empirical evidence may also impede the proper consideration of this question. In Germany, Weichselbaumer (2020) demonstrated labour market discrimination against Muslim women, by means of a correspondence test in which she combined fictitious CVs with photographs of the same woman, with or without a headscarf, and with a German and a Turkish name. She was thus able to show that compared to applicants with a German name, women with a Turkish name are less likely to be invited for an interview when applying for a job as an office worker, and the level of discrimination increases substantially if the applicant wears a headscarf.

3. Methodology

In this study, we investigate the combined effect of language use, ethnicity and a marker of religious identity on students' evaluation of university instructors. In order to do so, we developed a contextualised speaker evaluation experiment situated in the domain of higher education, with a 2 x 3 between-subjects design in which six groups of undergraduate students each evaluated a single instructor, presented to them in different guises. In this section we elaborate on the verbal and visual input (section 3.1), the participants (section 3.2), our procedure (section 3.3) and the measurements and analysis (sections 3.4 and 3.5).

3.1. Stimuli

3.1.1. Auditory input

For our speech samples, we selected a topic our respondents presumably would not have emotionally charged opinions about, viz. the Franco-Spanish war (1635-1659). The script for the lecture on this topic was based on the transcript of a real class taught at the history department of a Flemish university. The script was rendered in two versions which were syntactically identical but differed in the use of a number of morphological and phonological variables. In one version the speaker consistently used the Standard Dutch variants of those variables; in the other version she systematically used nonstandard variants, which are considered typical for Colloquial Belgian Dutch (Geeraerts and Van de Velde, 2013; see Appendix 1 for a full transcript of the auditory stimuli and Appendix 2 for an overview of the variables and their variants). Both versions contained 420 words and lasted approximately 2 minutes and 30 seconds. They were recorded by a single female speaker, a professional actress who is a native speaker of Belgian Dutch, born and raised in the greater Antwerp area. In order to sound realistic, the speaker used a mild version of her native regional accent, in line with the finding from previous research that Flemish teachers typically have an accent that renders their regional origin identifiable, rather than approximating the strictest accent norm, as exemplified by TV and radio newsreaders (Grondelaers and Van Hout, 2011, p. 219). In both speech clips, our speaker is identifiable as a speaker from the city of Antwerp.

3.1.2. Visual input

Instructor ethnicity was operationalized through the combination of a name and a photograph. For the Flemish identity we opted for the first name 'Sofie', which was common for girls born in Flanders around 1985, and combined it with the last name 'Jacobs', the fourth most common surname in Flanders. For the Maghrebi identity, we used the Arabic equivalent 'Safia' for the first name and selected the name 'El Jattari' from a list of Arabic surnames occurring in Belgium. We circulated an online survey to check whether participants would be likely to identify the names we selected as Flemish and Maghrebi respectively. 140 respondents were asked to indicate on a scale from 1 to 7 how likely they thought it was that people with these names have Moroccan or Flemish parents ($t = 34.55$, $df = 278$, p-value < 0.001). The results showed that Sofie Jacobs is associated with Flemish (Median = 6) rather than with Moroccan parents (Median = 1, Wilcoxon rank sum test with continuity correction: $W = 19257$, $p < 0.000$ effect size $r = -12$). The opposite is true for the name Safia El Jattari, which was considered to be a name given by Moroccan (Median = 6) rather than by Flemish parents (Median = 2, $W = 914$, $p < 0.000$, $r = -1.122$).

We took two photographs of the same woman, one with and one without a hijab, to combine with the different scenarios. We used the same person for all scenarios to rule out the possible impact of attractiveness, charisma or age (Baert and Decuypere, 2014; Riniolo et al., 2006). For the fabrication of the photographs, we hired a female model of northern African ethnicity (30 years of age) who could pass equally well as Flemish and Maghrebi. In a pre-test this was verified empirically. We asked the same pool of respondents to rate on a scale from one to seven how likely they thought it was that the person in the photograph without the hijab had Moroccan (Median = 4) or Flemish parents (Median = 4) ($W = 2702.5$, $p = 0.146$, $r = -0.12$).

By combining the names and photographs, three fictitious female university instructors were created. The three identities tested are shown in Figure 1.

Sofie Jacobs | Safia El Jattari | Safia El Jattari

Figure 1: Identities tested

These three identities were combined with the two language scenarios in a total of six conditions (cf. Table 1).

3.2. Participants

Our participants were undergraduates from STEM departments at a Flemish university. Six groups of first- and second-year students of Industrial Sciences participated in this study, during the first fifteen minutes of a session of the course 'Signals and systems' or 'Physics'. 332 participants completed the survey, but after inspection for outliers, the data of only 314 respondents were eventually included in the analysis (see Table 1).

	Name	**Picture**	**Language scenario**	**Number of participants**
Condition 1	Sofie Jacobs	Bareheaded	Standard Dutch	58
Condition 2	Sofie Jacobs	Bareheaded	Colloquial Dutch	83
Condition 3	Safia El Jattari	Hijab	Standard Dutch	44
Condition 4	Safia El Jattari	Hijab	Colloquial Dutch	38
Condition 5	Safia El Jattari	Bareheaded	Standard Dutch	38
Condition 6	Safia El Jattari	Bareheaded	Colloquial Dutch	53

Table 1: Six conditions

As we conducted the experiment at the start of six classes and used the students that were attending these classes as participants, we were not able to equally distribute participants over conditions based on their

background characteristics (e.g. their age, gender or nationality). Across the six groups, respondents had an average age of 20.26 years (SD = 1.64; range = 18-35). Approximately 87% were male, 13% were female. 97% of our respondents had the Belgian nationality; 3% had a different nationality or a double nationality (including the Belgian nationality). 95% of our respondents had two parents with the Belgian nationality, while 5% had at least one parent with a different nationality or a double nationality. 92% of our respondents spoke Dutch at home, 4% French and another 4% had a different home language. Most of them lived in the western half of Flanders: 26% lived in West-Flanders, 60.5% in East-Flanders, 9% in Antwerp and 4.5% in Flemish-Brabant. 83% of our respondents had obtained a degree in general secondary education and 17% had obtained a degree in technical secondary education.

3.3. Procedure

The experiment was conducted by master's students, who gave instructions to each experimental group in an identical manner, presenting themselves as students from the Faculty of Arts and Philosophy. Each group was presented with a single condition. The speech sample was presented as a recording of part of a lecture on the Franco-Spanish war that had been taught to history students at a Flemish university. The speaker was introduced as the instructor ('docent') of that course and it was announced that also the PowerPoint presentation that she had shown during her lecture would be shown on screen. The respondents were told that they would afterwards receive questions about the instructor. They were informed that no content questions would be asked, but that they nevertheless should try to listen carefully.
A photograph representing the instructor, accompanied by her name, was projected on the left half of a screen while the participants listened to the lecture. On the right half of the screen, the PowerPoint presentation was projected, in which the main points of the lecture appeared as the lecture progressed (see Figure 2).

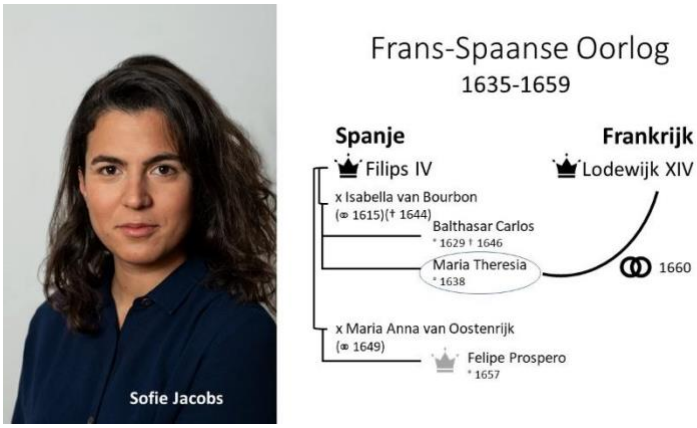

Figure 2: Projection of photograph, name and PowerPoint presentation (condition 1)

Immediately after hearing the lecture, participants were asked to complete an online questionnaire designed in LimeSurvey Version 2.73.1.

3.4. Measurements

Our questionnaire consisted of three phases. In the first phase, participants responded to 25 Likert statements on a 7-point scale ranging from 'completely disagree' to 'completely agree'. All scales were presented with the same directionality, going from negative to positive. The 25 items were chosen on the basis of previous studies, where they were included in the evaluative dimensions *competence*, *status*, *solidarity*, *dynamism*, *credibility* and *reported comprehensibility* (Ghyselen, 2009; Grondelaers and Van Gent, 2019; Impe and Speelman, 2007; Vandekerckhove and Cuvelier, 2007; Grondelaers and Van Hout, 2010; Grondelaers and Speelman, 2013, Lybaert, et al., 2020; Neuliep and Speten-Hansen, 2013). We

adapted them to the educational context and supplemented them with a number of statements that we selected ourselves. None of the Likert statements directly referred to the language variety used by the instructor. Most of them can therefore be considered indirect attitude measures. The items probing for comprehensibility ('I think this instructor speaks clearly', 'I find this instructor easy to understand', 'I have understood the explanation well', 'I think this instructor explains the matter clearly') can be considered semi-direct, since they could be taken to refer to the instructor's speech style.

The second phase consisted of 12 semantic differential items, with polar opposite descriptions at either end of 7-point scales (negative at the left end of the scale, positive at the right end). The semantic differential items contained opposite statements about the instructor's language and appearance (see Appendix 3 for an overview). The items on language use can be considered direct measures of language attitudes. In addition, participants were asked to indicate in which country they thought the instructor was born ('Belgium' or 'other'), and to specify what country if they thought she was born outside of Belgium. They were also asked to indicate in which of the five Flemish provinces they thought she lived.

Finally, the questionnaire contained a number of questions eliciting demographic information about the respondents, viz. their gender, year of birth, place of origin, degree, their own and their parents' nationality and home language(s).

The questionnaire was presented to the respondents on three consecutive web pages: Likert statements (in random order) on page 1, semantic differential items (in random order) and information about the speaker's provenance on page 2 and demographic information on page 3. Once participants loaded a new page, they were not able to return to the previous page. At the end of the survey, participants were asked to declare their informed consent.

3.5. Analysis

We conducted two primary analyses: a principal component analysis (PCA) and a multivariate analysis of variance (MANCOVA). All analyses were conducted in SPSS version 25.

The PCA served as an exploratory analysis and was used to reduce the dimensionality in the data by determining which items from the survey loaded onto the same components. Since the structure of the indirect and semi-direct items differed from that of the direct items (Likert scale structure vs. semantic differentials), the two parts of the survey were included in two separate PCAs. Prior to conducting the PCA, Cronbach's Alpha was checked to determine whether omitting an item would benefit the reliability of the survey. This was not the case for the indirect Likert statements (n = 24, α = 0.927). After the removal of two direct semantic differential items ('The instructor does not have vs. has a foreign accent' and 'The instructor does not have vs. has a regional accent'), the reliability of that section increased substantially (before removal: n = 12, α = 0.65; after removal: n = 10, α = 0.81).

For the PCA of the indirect and semi-direct items (Kaiser-Mayer-Olkin: 0.908; Bartlett's Test of Sphericity ($\chi^2(276) = 4321.53$, $p < 0.000$); det. > 0.00001) a four-factor solution was chosen, based on the eigenvalues (>1) and a visual examination of the scree plot. Jointly, these four factors explained 60.07% of the variance. Table 2 shows the component matrix of these four dimensions after Varimax rotation with Kaiser Normalization.

	Component			
	1	**2**	**3**	**4**
	Social attractiveness	**Professionalism**	**Comprehensibility**	**Authority**
I find this instructor entertaining.	.742			

I think this instructor has a warm personality	.724			
I think this instructor is popular with students.	.711			
I think this instructor is nice.	.681			
I enjoy being taught by this person.	.653			
I think this instructor is helpful.	.650			
I find this instructor cool.	.635			
I would like to be taught by this person.	.634			
I think this instructor can be funny.	.563			
I think students look up to this instructor.	.552			
I think this instructor knows her field.		.827		
I think this instructor is reliable.		.771		
I think this instructor is intelligent.		.683		
I think this instructor has a successful academic career.		.614		
I think this instructor is credible.		.589		
I think this instructor has a lot of professional experience.		.568		
I find this instructor convincing.		.527		
I think this instructor speaks clearly.			.825	
I find this instructor easy to understand.			.804	
I have understood the explanation of this instructor well.			.705	
I think this instructor explains the matter clearly.			.683	
I think this instructor is assertive.				.579
I think this instructor has authority.				.555
This instructor comes across as self-confident.				.476

Table 2: PCA 1 – PCA Likert statements (loadings < 0.4 omitted for readability purposes)

Labelling the components is not a straightforward matter, given the considerable variability in the labels used in the existing literature. Component 1 comprises features that in previous studies have been labelled as 'Social attractiveness' (*funny*, *entertaining*; cf. Impe and Speelman, 2007; Lambert et al., 1966), 'Dynamism' (*nice*, *cool*, *popular*, *entertaining*; cf. Grondelaers and Kristiansen, 2013; Grondelaers and Speelman, 2013), 'Solidarity' (*nice*, *popular*, *entertaining*; cf. Ghyselen, 2009; Bayard et al., 2001) and 'Personal integrity' (*helpful*; cf. Impe and Speelman, 2007; Lambert et al., 1966). We chose to use the most encompassing label and named the dimension 'Social attractiveness'. Some of the features comprised in component 2, notably *intelligent*, *successful*, *reliable* and *experienced*, have likewise received various labels in previous studies: 'Competence' (Lambert et al., 1966; Bayard et al., 2001), 'Status' (Impe and Speelman, 2007; Ryan & Carranza, 1975) or 'Superiority' (Grondelaers and Kristiansen, 2013; Zahn and Hopper, 1985). *Reliable* has in previous studies also been captured by the label 'Integrity' (Grondelaers and Speelman, 2013), *credible* by 'Credibility' (Neuliep and Speten-Hansen, 2013). Given that all features refer to the instructor's professional competence, we chose for the encompassing label 'Professionalism'. Component 3 can straightforwardly be captured by the label 'Comprehensibility'. Component 4 comprises two features that have previously been labelled 'Dynamism', viz. *assertive* (Grondelaers and Speelman, 2013) and *self-confident* (Grondelaers and Kristiansen, 2013; Zahn and Hopper, 1985). However, given that component 1 also contains features that have been labelled as 'Dynamism', we chose to label component 4 'Authority'.

The PCA (Varimax rotation with Kaiser Normalization) of the indirect items was first run using a three-component solution explaining 73% of the variance (KMO: 0.826, Bartlett's Test of Sphericity ($X^2(45)$ = 2013, $p < 0.000$) / det. > 0.002). Since the third component only yielded one loading ("This instructor is not a native speaker of Dutch. | This instructor is a native speaker of Dutch"), we ran the PCA again without that variable (KMO: 0.813, Bartlett's Test of Sphericity ($X^2(36) = 1697$, $p < 0.000$) / det. = 0.004), resulting in a two-component solution (reliability $n = 9$, $\alpha = 0.789$) (see Table 3).

	Component	
	5	6
	Physical attractiveness	Standardness
I don't think this instructor is beautiful. | I think this instructor is beautiful.	.799	
I don't think this instructor is pretty. | I think this instructor is pretty.	.789	
I think this instructor isn't good-looking. | I think this instructor is good-looking.	.787	
I don't find this instructor attractive. | I find this instructor attractive.	.713	
This instructor speaks sloppy. | This instructor speaks properly.		.649
This instructor does not speak beautifully. | This instructor speaks beautifully.		.601
This instructor does not speak correct Dutch. | This instructor speaks correct Dutch.		.666
This instructor speaks a dialect. | This instructor speaks Standard Dutch.		.513
This teacher doesn't speak proper Dutch. | This teacher speaks proper Dutch		.523

Table 3: PCA 2 - PCA semantic differential items

Component 5, comprising all judgements of the speaker's looks, was labelled 'Physical attractiveness' (as opposed to 'Social attractiveness'). Component 6 comprises judgements of the beauty, properness, correctness and standardness of instructor's language use and can therefore be labelled 'Standardness' (cf. Grondelaers and Kristiansen, 2013, p. 17; Lybaert, 2017).

In order to determine the impact of the conditions while controlling for potential covariates, we conducted a MANCOVA using the PCA components as dependent variables. To calculate these variables, we used weighed items scores, meaning that the original scores for each item were multiplied by the factor loading associated with that item. All weighed item scores belonging to the same component were then averaged to form the outcome variable. Prior to conducting the MANCOVA, the data were checked for outliers. We identified outliers by visually examining the boxplots of the dependent variables and by excluding participants with values outside of the following range: ((3rd quartile + 1.5*interquartile range), (1st quartile – 1.5*interquartile range)). The responses of 18 participants were removed, while respecting the proportionality of the dataset (i.e., ensuring that removing the outliers did not disturb the balance between the conditions), resulting in a sample size of $n = 314$. A post-hoc power analysis conducted using G*power version 3.1.9.7 (Faul et al., 2009) showed that at $\alpha = 0.05$, the sample size was adequate (power = 0.999) to detect medium effects ($f^2 = 0.15$). The model met all assumptions (Levene's test, $p > 0.05$; Cook's $d < 0.08$; visual inspection of linearity of residuals). We operationalized five covariates we assumed might impact judgments due to perceived affinity with the speaker: gender ($p < .000$; : $\eta_p^2 = .081$), age ($p = .82$; : $\eta_p^2 = .01$), province ($p = .19$; : $\eta_p^2 = .029$), nationality ($p = .001$; : $\eta_p^2 = .074$), L1 ($p < .204$; : $\eta_p^2 = .028$). Two covariates (gender, nationality) were retained, since they were found to have a significant medium effect on the multivariate model.

4. Results

4.1. Descriptive data

Table 4 shows the minimum, maximum and mean score and standard deviation for the six components (of which all items were evaluated on 7-point-scales) across the conditions; Table 5 shows the mean score,

standard error (SE) and standard deviation (SD) per condition for the six components. The items in the components 'Professionalism', 'Authority' and 'Social Attractiveness' contained indirect Likert statements (with a 7-point scale ranging from 'completely disagree' to 'completely agree'). While for the component 'Authority' the mean score per condition is the most negative (ranging between 2.66 and 2.85 out of 7), the mean score is a bit less negative for 'Social attractiveness' (mean ranging between 3.17 and 3.35) and 'Professionalism' (mean ranging between 3.32 and 3.67). 'Comprehensibility', the component containing semi-direct Likert statements, received more neutral mean scores (ranging between 3.86 and 4.65). The same goes for the semantic differential items in the direct components 'Physical attractiveness' (mean ranging between 3.49 and 4.63) and 'Standardness' (mean ranging between 3.62 and 4.82). The standard deviations (see Table 4 and 5) are rather high (>10% of mean) for the variables 'Physical attractiveness', 'Comprehensibility' and 'Standardness'. Since the variables we modified had to do with pronunciation and physical appearance, large standard deviations were to be expected.

	Minimum	**Maximum**	**Mean**	**Standard Deviation (SD)**
Social attractiveness	1,97	4,58	3,2696	0,49965
Professionalism	2,18	4,52	3,4210	0,45679
Comprehensibility	2,06	5,28	4,2478	0,67521
Authority	1,80	3,76	2,7643	0,38721
Physical attractiveness	1,59	6,32	4,1220	0,98072
Standardness	1,41	5,40	3,8910	0,81537

Table 4: Minimum, maximum, mean and standard deviation across conditions per component

	CONDITION 1 (n=58)		**CONDITION 2 (n=83)**		**CONDITION 3 (n=44)**		**CONDITION 4 (n=38)**		**CONDITION 5 (n=38)**		**CONDITION 6 (n=53)**	
	Mean (SE)	**SD**	**Mean (SE)**	**SD**	**Mean (SE)**	**SD**	**Mean (SE)**	**SD**	**Mean (SE)**	**SD**	**Mean (SE)**	**SD**
Social attractiveness	3.17 (0.06)	0.46	3.27 (0.06)	0.55	3.26 (0.07)	0.48	3.35 (0.10)	0.59	3.23 (0.07)	0.43	3.35 (0.06)	0.44
Professionalism	3.32 (0.06)	0.44	3.41 (0.05)	0.47	3.67 (0.06)	0.37	3.37 (0.08)	0.50	3.36 (0.07)	0.43	3.41 (0.06)	0.45
Comprehensibility	3.86 (0.01)	0.74	4.24 (0.07)	0.62	4.65 (0.07)	0.48	4.2 (0.13)	0.78	4.54 (0.07)	0.46	4.16 (0.09)	0.62
Authority	2.66 (0.05)	0.39	2.74 (0.04)	0.37	2.82 (0.06)	0.39	2.76 (0.06)	0.36	2.79 (0.07)	0.44	2.85 (0.05)	0.38
Physical attractiveness	4.63 (0.12)	0.9	4.27 (0.09)	0.84	3.72 (0.16)	1.06	3.49 (0.13)	0.79	4.24 (0.16)	0.98	4.03 (0.14)	0.98
Standardness	4.27 (0.08)	0.65	3.71 (0.09)	0.88	4.82 (0.09)	0.621	3.62 (0.15)	0.905	4.41 (0.13)	0.79	3.85 (0.12)	0.89

Table 5: Mean, standard error and standard deviation per condition and component

We also asked our respondents in which country they believed the instructor was born and in which province they believed she lived. On average, 86.6% of the respondents believed the instructor was born in Belgium. 13.4% of the respondents believed she was foreign-born: Morocco, the Middle East and Spain were mentioned most often as country of origin. As far as her regional origin is concerned, the instructor was located mostly in East-Flanders (43%), followed by the province of Antwerp (37.9%). In the standard guises (conditions 1, 3 and 5) the instructor was located most often in East-Flanders (48.6%), while in the colloquial guises (2, 4 and 6) most respondents believed she lived in Antwerp (44.8%).

As discussed above, two semantic differential items were excluded from the multivariate analysis. A Wilcoxon rank sum test was computed to assess whether there were significant differences in perceived foreign accent and perceived regional accent between scenarios. We computed effect size *r* to determine the magnitude of the difference (see Table 6). In general the respondents did not perceive a strong foreign accent. The Median for the differential 'The instructor does not have vs. has a foreign accent' was 2 in all conditions except for condition 2 (Flemish name, Colloquial Dutch) and we only observed a significant difference with a small effect between conditions 2 and 6: in condition 6 (Arabic name, no hijab, Colloquial Dutch) less of a foreign accent was perceived than in condition 2 (Flemish name, Colloquial Dutch). Concerning the differential 'The instructor does not have vs. has a regional accent', there were more significant differences between conditions, and the effect sizes were generally larger (median effect size *r* of significant results = -0.48). In general, the regional accent of the Maghrebi instructor (both with and without headscarf) was perceived as stronger, and this was especially the case in the colloquial conditions, which were perceived to have a stronger regional accent than their Standard Dutch counterparts.

	Condition	**Condition**	***W***	***p***	**Effect size *r***
The instructor does not have vs. has a foreign accent	2 (Med = 3)	6 (Med = 2)	2664.5	0.03	-0.182
The instructor does not have vs. has a regional accent	1 (Med = 4)	2 (Med = 5)	1322	<0.001	-0.389
		3 (Med = 3)	1626	0.02	-0.238
		4 (Med = 6)	376.5	<0.001	-0.564
		6 (Med = 6)	657	<0.001	-0.501
	2	3 (Med = 3)	2891	<0.001	-0.485
		4 (Med = 6)	1226	0.04	-0.183
		5 (Med = 4)	2249.5	<0.001	-0.110
	3	4 (Med = 6)	206.5	<0.001	-0.654
		6 (Med = 6)	363	<0.001	-0.598
	4	5 (Med = 4)	1176	<0.001	-0.551
	5	6 (Med = 6)	450.5	<0.001	-0.477

Table 6: Wilcoxon Rank Sum Test (effect size R) on the items foreign accent and regional accent

4.2. Multivariate analysis

As Table 7 shows, using Pillai's Trace, we found a significant, medium effect of condition on the outcome variables, after controlling for gender and nationality ($F(30, 1515) = 6.953$, $p < 0.000$).

	Value	***F***	**Hypothesis df**	**Error df**	**Sig.**	η_p^2	**Obs. Power**
Intercept	0.688	110.008	6	299	.000	0.688	1
Gender	0.085	4.629	6	299	.000	0.085	0.988
Nationality	0.07	3.769	6	299	.001	0.07	0.962
Condition	0.605	6.953	30	1515	.000	0.121	1

Table 7: Multivariate test (Note: Test statistic: Pillai's Trace)

Tests of between-subject effects (Table 8) show that the condition has a significant impact on the outcome variables 'Professionalism' (η_p^2 = 0.06, R^2_{adj} = .057), 'Comprehensibility' (η_p^2 = .14, R^2_{adj} = 0.119),

'Physical attractiveness' (η_p^2 = .13, R^2_{adj} = 0.1327), and 'Standardness' (η_p^2 = .213, R^2_{adj} = 0.214). No significant or substantial effect were found for 'Social attractiveness' and 'Authority'.

	Type III Sum of Squares	**df**	**Mn Sq.**	***F***	**Sig.**	η_p^2	**Obs Power**
Social attractiveness	1.756	5	0.351	1.513	.185	0.024	0.529
Professionalism	3.842	5	0.768	3.9	.002	0.06	0.943
Comprehensibility	19.478	5	3.896	9.695	.000	0.138	1
Authority	1.282	5	0.256	1.745	.124	0.028	0.599
Physical attractiveness	37.473	5	7.495	8.948	.000	0.128	1
Standardness	42.539	5	8.508	16.415	.000	0.213	1

Table 8: Between-subject effects

Figures 3 to 6 show the estimated marginal means for the four outcome variables on which the condition has a significant effect (see also Appendix 4 for estimated marginal means comparison). Figure 3 shows that the standard-speaking Maghrebi instructor with hijab (condition 3) was perceived as significantly more professional than her colloquial-speaking counterpart (condition 4) and all other guises. No other significant differences were observed.

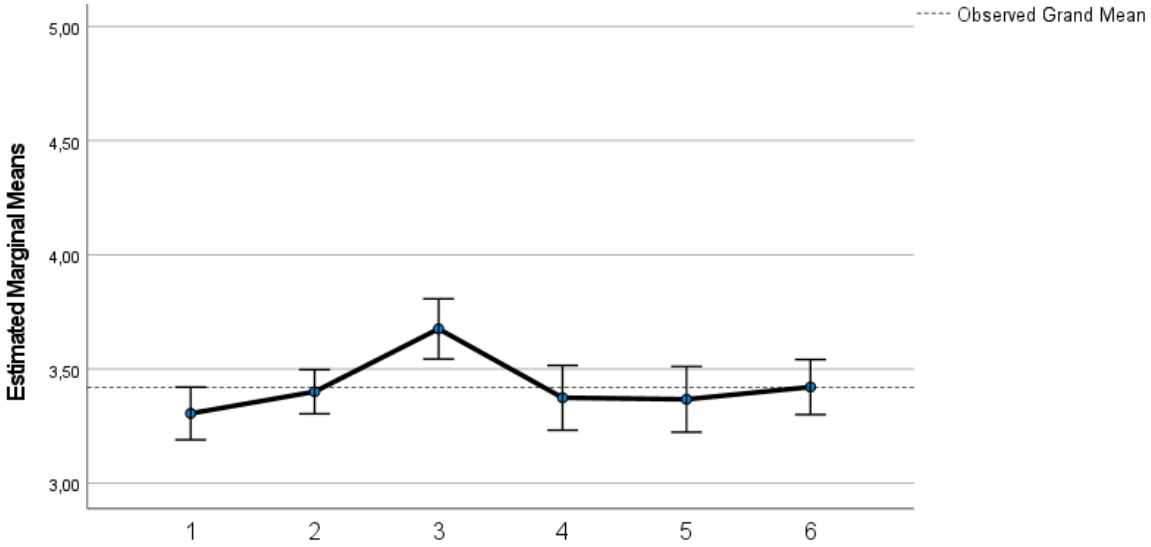

Figure 3: Estimated marginal means for 'Professionalism' (Note. Error bars: +/- 2 SE)

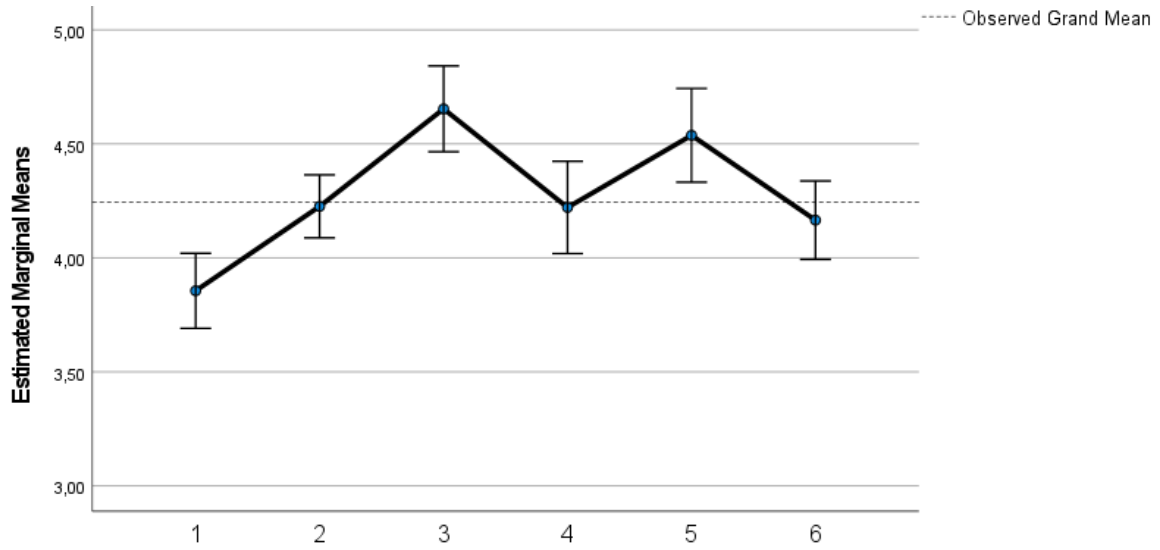

Figure 4 shows that 'Comprehensibility' scores were significantly lower in condition 1 (Flemish name, no hijab, Standard Dutch) than in all other conditions. Conditions 3 and 5 received the highest scores on 'Comprehensibility', which means that the exact same Standard Dutch audio combined with a Maghrebi speaker received significantly higher ratings than combined with a Flemish speaker. 'Comprehensibility' scores in condition 3 (Arabic name, hijab, Standard Dutch) were also significantly higher than those in conditions 4 and 6 (Arabic name, Colloquial Dutch). Condition 5 (Arabic name, no hijab, Standard Dutch) received significantly higher 'Comprehensibility' scores than condition 6 (Arabic name, no hijab, Colloquial Dutch). Thus we found significant differences for 'Comprehensibility' between the standard and colloquial guises in all three identities, but whereas the standard-speaking Flemish instructor is perceived as significantly less comprehensible than her colloquial-speaking counterpart, the effect is reversed for the Maghrebi instructor, both with and without hijab.

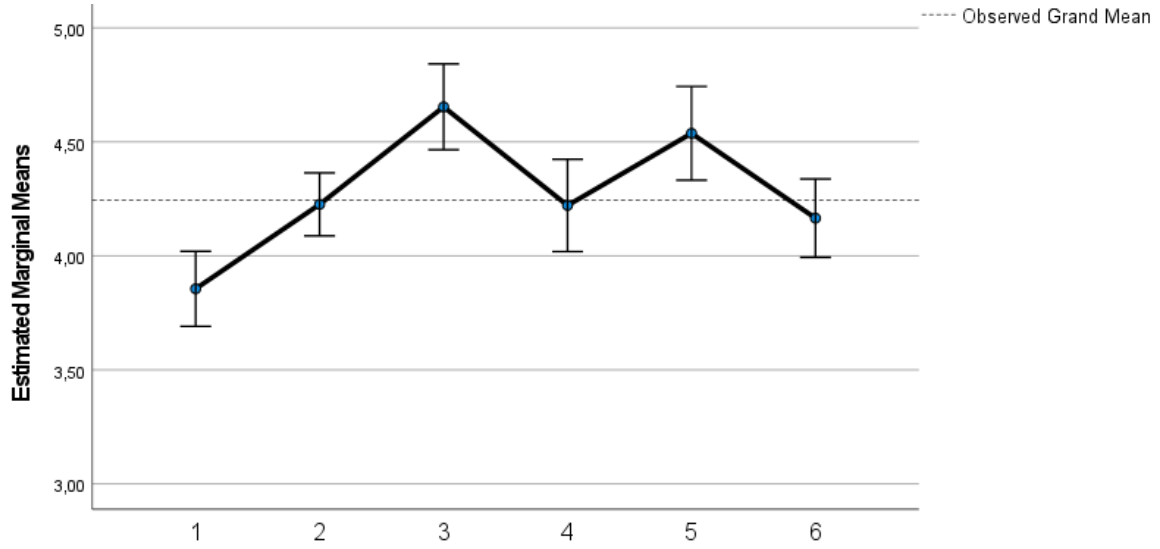

Figure 4: Estimated Marginal Means for 'Comprehensibility' (Note. Error bars: +/- 2 SE)

As Figure 5 shows, condition 1 was judged as physically most attractive: the standard-speaking Flemish instructor received significantly higher scores than all other conditions except for 5 (Arabic name, no hijab, Standard Dutch). Condition 2 (Flemish name, Colloquial Dutch) was judged significantly more attractive

than both conditions with hijab, viz. 3 and 4, which also both scored significantly lower than condition 5 (Arabic name, Standard Dutch, no hijab). Condition 4 (Arabic name, hijab, Colloquial Dutch) in addition scored significantly lower than her unveiled counterpart in condition 6. Thus, the conditions featuring a hijab showed a significant drop in 'Physical attractiveness' scores. No significant differences could be found between the standard and colloquial guises within both Maghrebi identities.

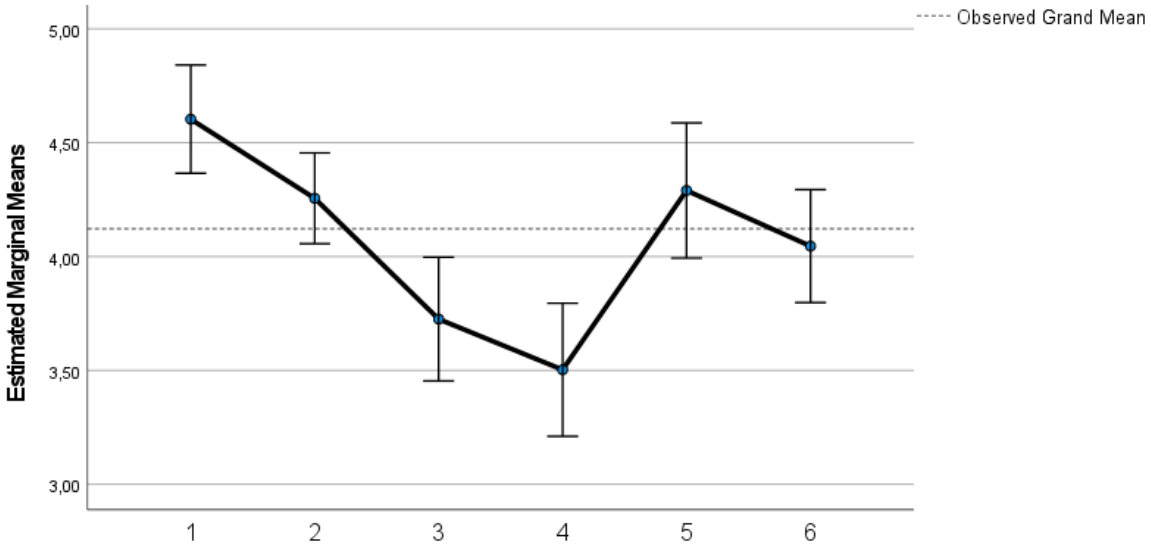

Figure 5: Estimated Marginal Means for 'Physical attractiveness' (Note. Error bars: +/- 2 SE)

Lastly, Figure 6 shows that all standard guises (1, 3 and 5) were judged as significantly more standard than their colloquial counterparts (2, 4 and 6). The highest scores were awarded in condition 3, featuring the hijab, Arabic name and standard language use. This condition received significantly higher scores than all other Maghrebi guises (conditions 4, 5 and 6). No significant differences in 'Standardness' were found among the colloquial guises.

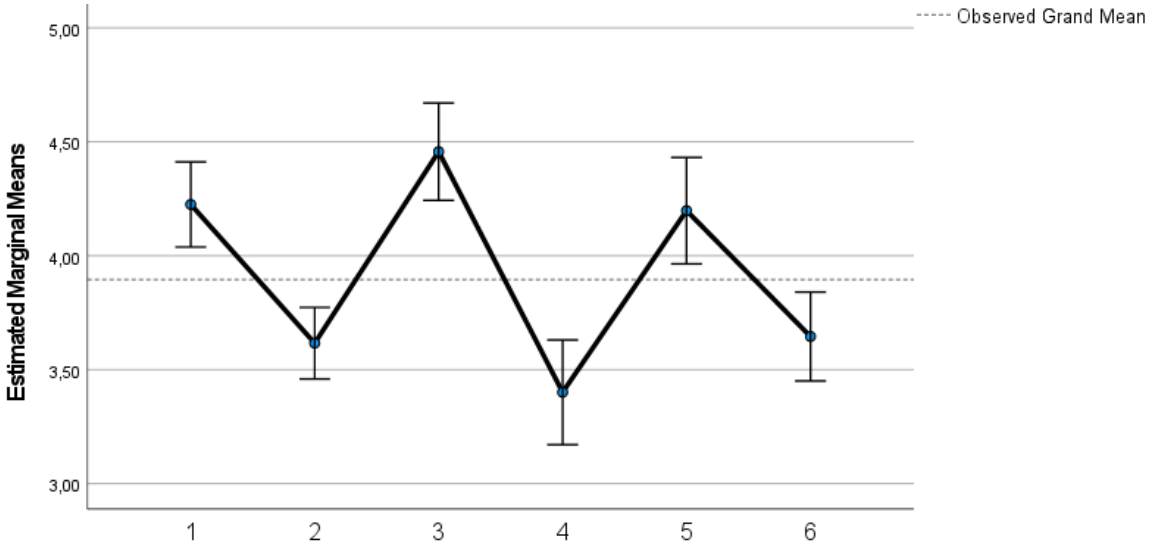

Figure 6: Estimated Marginal Means for 'Standardness' (Note. Error bars: +/- 2 SE)

5. Discussion

The results of our speaker evaluation experiment show that the instructor was evaluated differently depending on her language use and on the ethnic identity ascribed to her. If we take a closer look at the results, we observe a discrepancy between the indirect measures on the one hand and the semi-direct and direct measures on the other hand: the indirect measures returned no or small significant effects between conditions, while the semi-direct and direct measures returned more and larger effects.

As for the indirect measures, we observed that the impact of condition corrected for gender was not significant for 'Social attractiveness' and 'Authority': the standard guises were perceived as no less socially attractive than the colloquial guises and the Maghrebi identities were deemed no less socially attractive than the Flemish identity. Similarly, the colloquial guises were judged no less authoritative than the standard guises, nor did the speaker's ethnic identity impact on this dimension. We did observe a small effect for 'Professionalism', where only the standard-speaking instructor with hijab stood out as significantly more professional than her colloquial-speaking counterpart, the Flemish instructor and the Maghrebi instructor without hijab. However, in two out of three identities no significant differences could be observed between the standard and colloquial guises for 'Professionalism'.

Important to mention, moreover, is that the scores for the indirect measures were situated on the negative side of the 7-point Likert scale: in all conditions the instructor was downgraded for 'Authority' and (slightly less so) for 'Professionalism' and 'Social attractiveness'. The lack of positive scores for the standard guises might be the result of our deliberate choice not to work with speech clips containing 'unaccented' Belgian Standard Dutch, as represented by TV and radio newsreaders. A 'newsreader accent', devoid of any regional marking, is usually found to score highest for prestige when included in speaker evaluation designs (Grondelaers, 2013). However, we assumed that such an accent would come across as unrealistic in this particular context. We deliberately opted for Standard Dutch with a mild regional accent, given that previous research found that 'Teacher Dutch' in Flanders is generally marked by a regional accent (Grondelaers et al., 2011; Delarue, 2013). Another explanation for the general downgrading might be that our instructor did not speak with an East-Flemish accent, which would be endogenous for 60,5% of our respondents, but with an exogenous Antwerp accent, a choice which might have impacted ratings negatively too. However, the instructor's accent was not recognized as Antwerpian by the majority of our respondents (37.9% of our respondents believed the instructor lived in Antwerp, while 43% located her in East-Flanders). A final explanation for our instructor's downgrading, and perhaps the most likely one, might be the subject matter: science students may have little affinity with, and perhaps even a general adversity towards, a history topic, which is entirely outside their domain of study. In order to substantiate this hypothesis, it would be worth exploring if a population of humanities (or more specifically history) students would evaluate the instructor more positively. Apart from the negative to neutral mean scores, however, it remains surprising that the colloquial guises were generally not downgraded *more* for 'Authority' and 'Professionalism' compared to the standard guises, especially if we take into account that the formal educational context in which our speaker was situated is traditionally associated with Standard Dutch. Thus, our results do not corroborate the assumption that Standard Dutch is a *conditio sine qua non* for instructors to be perceived as professional and authoritative in the context of higher education.

In contrast to the general pattern of downgrading on the indirect measures, average scores on the semi-direct and the direct measures were for all conditions located more around the middle of the scale, and we did observe large effects for 'Physical attractiveness', 'Standardness' and 'Comprehensibility'. The Flemish identity was perceived as physically more attractive than the Maghrebi ones, especially the one with hijab. In terms of 'Standardness', Standard Dutch always scored better than Colloquial Dutch, and here we observed that, similar to the scores on 'Professionalism', the standard-speaking Maghrebi instructor with hijab stood out positively. For 'Comprehensibility', we also observed that Standard Dutch generally

performed better than Colloquial Dutch, except – surprisingly – for the Flemish identity, where Standard Dutch received the lowest of all scores.

So what do the results of our study tell us about the position of Belgian Standard Dutch and Colloquial Dutch in higher education, and about the vitality of the standard language ideology? In the direct and semi-direct measures we clearly observe a pattern in line with the traditional standard language ideology: Standard Dutch generally performed better for 'Standardness' and 'Comprehensibility'. Related to that, the semi-direct and direct measures also seem to indicate that a Maghrebi instructor – especially when wearing a hijab – can distinguish herself linguistically by speaking Standard Dutch (and far less when speaking Colloquial Dutch). The 'bonus' for 'Professionalism' and 'Comprehensibility' that the veiled, and thus probably most 'un-Flemish' looking speaker receives for speaking Standard Dutch might be interpreted as a sign that participants' perceptions are in line with the dominant discourse that a good competence of Dutch is a necessary precondition for foreigners' and ethnic minority members' inclusion and social mobility in Flemish society, and testifies to their 'willingness to integrate' (Jaspers, 2012; 2013; Van Hoof et al., 2020). These results can thus be interpreted as signs of the vitality of the standard language ideology. In general, however, the results for the indirect measures are not unambiguously in line with this ideology. Standard Dutch does *not* perform significantly better than Colloquial Dutch on the measures 'Authority' and 'Professionalism', which may indicate that Colloquial Dutch is equally acceptable in the context of higher education. Our results suggest that speech standardness only has a limited impact on *in situ* judgements of an instructor's professional competences. It seems that at the subconscious level, standard language ideological beliefs and stereotypical ideas about language are less strongly expressed in this contextualized design, compared to decontextualized experiments. This is not surprising, as in decontextualized experiments, respondents can only fall back on the speaker's voice and accent to judge them, and may therefore be more readily inclined to base their judgements on linguistic stereotypes.

As for the impact of ethnicity and wearing a hijab on our respondents' perceptions and attitudes, we did not observe general downgrading for the Maghrebi identities compared to the Flemish identity. While wearing a hijab led to downgrading of the instructor in terms of 'Physical attractiveness', the standard-speaking Maghrebi instructor with hijab was perceived as significantly more professional than all other guises, and both standard-speaking Maghrebi identities – with and without hijab – were evaluated as significantly more comprehensible and more standard than their colloquial counterparts on the one hand and the Flemish identity on the other. The Maghrebi instructor was not downgraded any more than all other guises on 'Authority' and 'Social attractiveness' either. Moreover, the respondents did not report hearing highly foreign-accented speech in any of the conditions, and foreign accent was not perceived as stronger in the Maghrebi identities.

In contrast to the results of Rubin and colleagues (Rubin, 1992; Rubin et al., 1999; Rubin and Smith, 1990), we did not find evidence of negative 'reverse linguistic stereotyping' (Kang and Rubin, 2009) vis-à-vis lecturers with an ethnic minority background: the Maghrebi lecturers in our study were *not* perceived as less intelligible, having poorer teaching skills and having a stronger foreign accent than the majority identity lecturer. Our results also deviate from the findings of Grondelaers and Van Gent (2019), in whose speaker evaluation experiment a male Moroccan identity (indexed by Arabic names) was downgraded on superiority (with measures such as 'this person got good grades', 'has a lot of professional experience' and 'has good management skills') and upgraded on what Grondelaers and Van Gent (2019) called 'macho' dynamism (i.e. the traits 'hip', 'tough' and 'cool'), even when no Moroccan accent could be discerned. Obviously, the design of both studies differed: the speakers of Grondelaers and Van Gent (2019) were men, and ethnic identification was possible only through first name representation. In addition, the speakers of Grondelaers and Van Gent (2019) were presented as applicants of whom the researchers wanted to map out to potential employers what image their voice evokes. In our contextualized experiment, by contrast, respondents assessed a speaker who they knew was a history instructor at a Flemish university, which implies that the speaker is a highly educated person who is already doing well professionally.

Apart from these differences, the question still arises why our veiled standard-speaking Maghrebi instructor was not downgraded, and even upgraded (or less fiercely downgraded than the other conditions) on 'Professionalism' and 'Standardness'. This may be related to our respondents' expectations towards the instructor. 'Language Expectancy Theory' (Burgoon and Miller, 1985; Burgoon, Denning & Roberts, 2002) states that we all have linguistic expectations when we communicate with someone, and expectancy violations may have a positive or negative impact on how we perceive that person. Our respondents may not have expected a veiled Maghrebi instructor to be able to communicate fluently in Belgian Standard Dutch. The instructor may therefore have positively broken their expectations, resulting in positive evaluations. This positive breach of expectations is perhaps broader than language alone: the idea that a woman with a migrant background and a headscarf makes a career as a university instructor in Belgium may exceed the expectations of our respondents. It is also not unlikely that we are witnessing the effect of a social desirability bias: the respondents were aware they were engaging in a rating activity, and out of a concern of not appearing racist, respondents of the conditions with the Maghrebi instructor (especially where she is wearing a hijab) may have underreported their socially undesirable attitudes towards Maghrebi women, and even compensated them for the competence measures. The more positive scores on the direct and semi-direct questions indeed point in the direction of social desirability (cf. Pantos and Perkins, 2012, who observed a bias in favour of foreign accented speech when directly measured, but not indirectly, which supported the argument that directly measured attitudes are cognitively controllable). However, a social desirability bias would also lead us to expect similar effects for both veiled identities, i.e. also an upgrading of the colloquial guise with hijab. This difference between the two veiled guises cannot be explained by a social desirability bias. A question at the end of the study gauging whether participants guessed the aim of the study would have provided interesting insights in that respect.

6. Conclusion

In this paper we have reported a contextualised speaker evaluation experiment probing into the combined effect of language variation (standard vs. vernacular varieties), ethnic identity (Flemish vs. Moroccan) and wearing a headscarf on students' evaluation of a female university instructor in Flanders. We were particularly interested in the potential impact of adding context to a controlled speaker evaluation experiment, as in our everyday lives we hardly ever judge speakers regardless of the specific context in which they operate. We incorporated as much context as possible, by providing respondents with the speakers' name and appearance (and thus information about the speakers' gender, ethnicity and age) and by creating a specific educational setting.

Both in terms of the effect of language variation and the effect of ethnicity and a headscarf, our contextualized study yielded somewhat surprising results. Contrary to what the traditional standard language ideology would lead us to expect, we observed a much less pronounced effect of the choice for Standard or Colloquial Dutch on our respondents' attitudes than in previous decontextualized attitude studies. Compared to Colloquial Dutch, Standard Dutch was not upgraded on classic status dimensions, which remains surprising in view of the formal educational context we created in our experiment, a context pre-eminently associated with Standard Dutch. This suggests that the impact of the standard language ideology, which our direct measure did show clear traces of, may be less palpable when informants assess speakers on indirect measures and *in situ*, rather than *in abstracto*. In addition, and contrary to what one might expect on the basis of the available evidence on ethnicity-based discrimination in Flanders, we did not witness any general downgrading of the ethnic minority identity and the headscarf: although there was a negative effect on perceived 'Physical attractiveness', there was no effect on other dimensions ('Authority', 'Social attractiveness') and on still others a positive effect, but only when combined with standard speech ('Comprehensibility', 'Professionalism').

Our experiment opens up several perspectives for future research. It would be interesting to incorporate foreign or ethnic minority accents in a similar design, or apply the design to other contexts perceived as

formal, in order to further explore the interplay of linguistic and non-linguistic cues in speaker assessment and gauge the impact of language ideologies on specific contexts. Finally, it is worth noting that our experiment has a real-life equivalent, in that most university instructors these days are subjected to teaching evaluations, the scores of which may have an impact on their academic careers. Whereas the literature on the validity of student evaluation of teaching in higher education has examined gender and racial bias fairly extensively (Chisadza et al., 2019; Mengel et al., 2019; Spooren et al. 2013), language as a factor affecting student evaluations has received considerably less attention (but see Subtirelu, 2015). Given that that literature moreover indicates that first impressions are formed very early on and are correlated with end-of-semester evaluations (Ambady & Rosenthal, 1993; Clayson, 2013; Laws et al., 2010), there is also a clear practical relevance to further experimentally assessing whether instructors' speech style and ethnicity have an impact on students' evaluations of the quality of their teaching.

Acknowledgements

We are grateful to Eline De Coninck, Debbie Gens and Marlies Lameire for their help in collecting the data for this study, to Dounia Bourabain and Pieter-Paul Verhaeghe for useful literature suggestions, and to the two anonymous reviewers for their helpful comments on an earlier version of this paper. All remaining errors are our own.

Funding

The authors received financial support from the host institution.

References

Agha, A., 2003. The social life of cultural value. Language and Communication 23, 231-273.

Ambady, N., Rosenthal, R., 1993. Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. Journal of Personality and Social Psychology 64, 431–441.

Baert, S., Decuypere, L., 2014. Better sexy than flexy? A lab experiment assessing the impact of perceived attractiveness and personality traits on hiring decisions. Applied Economics Letters 21 (9), 597-601.

Baert, S., Cockx, B., Gheyle, N., Vandamme, C., 2015. Is there less discrimination in occupations where recruitment is difficult? ILR Review 68, 467-500.

Baert, S., Vujić, S., 2016. Immigrant volunteering: A way out of labour market discrimination? Economics Letters 146, 95-98.

Baert, S., Albanese, A., du Gardein, S., Ovaere, J., Stappers, J., 2017. Does work experience mitigate discrimination? Economics Letters 155, 35-38.

Baert, S. 2018. Hiring Discrimination: An Overview of (Almost) All Correspondence Experiments since 2005. In: Gaddis, S. (Ed.), Audit Studies: Behind the Scenes with Theory, Method, and Nuance. Springer, New York, pp. 63-77.

Bayard, D., Weatherall, A., Gallois, C., Pittam, J., 2001. Pax Americana? Accent attitudinal evaluations in New Zealand, Australia and America. Journal of sociolinguistics 5, 22-49.

Burgoon, M., Miller, G., 1985. An expectancy interpretation of language and persuasion. In: Giles, G., Clair, R. (Eds.), The social and psychological contexts of language. Lawrence Erlbaum Associates, London, pp. 199-229.

Burgoon, M., Denning, V.P., Roberts, L., 2002. Language Expectancy Theory. In: Dillard, J.P, Pfau, M. (Eds.), The Persuasion Handbook. Developments in Theory and Practice. Sage, Thousand Oaks/London/New Delhi, pp. 117-136.

Clayson, D. E., 2013. Initial Impressions and the Student Evaluation of Teaching. Journal of Education for Business 88 (1), 26–35.

Coupland, N. 2010. Language, ideology, media and social change. In: Junod, K., Maillat, D. (Eds.), Performing the self. Gunter Narr, Tübingen, pp. 127-151.

Chisadza, C., Nicholls, N., Yitbarek, E., 2019. Race and gender biases in student evaluations of teachers. Economics Letters 179, 66-71.

Delarue, S., 2013. Teachers' Dutch in Flanders: the last guardians of the standard? In: Kristiansen, T., Grondelaers, S. (Eds.), Language (de)standardisation in Late Modern Europe: experimental studies. Novus, Oslo, pp. 193-226.

Deprez, K., 1981. Naar een eigen identiteit: resultaten en evaluatie van tien jaar taalsociologisch en sociolinguistisch onderzoek betreffende de standaardtaal in Vlaanderen. Catholic University of Leuven, unpublished dissertation.

Dragojevic, M., Berglund, C., Blauvelt, T., 2018. Figuring out who's who: The role of social categorization in the language attitudes process. Journal of Language and Social Psychology 37 (1), 28-50.

Eckert, P., 2008. Variation and the indexical field. Journal of Sociolinguistics 12 (4), 453-476.

Faul, F., Erdfelder, E., Buchner, A., Lang, A., 2009. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. Behavior Research Methods 41, 1149-1160.

Garrett, P., 2010. Attitudes to Language. Cambridge University Press, Cambridge.

Geeraerts, D., Van de Velde, H., 2013. Supra-regional characteristics of colloquial Dutch. In: Hinskens, F., Taeldeman, J. (Eds.), Language and Space. An International Handbook of Linguistic Variation. De Gruyter Mouton, Berlin, pp. 532-556.

Geerts, G., Nootens, J., van den Broeck, J., 1980. Opinies van Vlamingen over dialekt en standaardtaal. In: Geerts, G., Hagen, A. (Eds.), Sociolinguïstische studies 1. Bijdragen uit het Nederlandse taalgebied. Wolters-Noordhoff, Groningen, pp. 233-256.

Ghyselen, A., 2009. Ne zelfverzekerde leraar of gewoon nen enthousiaste mens? Een matched-guise onderzoek naar de attitude tegenover tussentaal bij West-Vlamingen. Taal en Tongval 61 (2), 83-113.

Grondelaers, S., 2013. Attitude measurements in the Low Countries. In: Hinskens, F., Taeldeman, J. (Eds.), Language and Space. An International Handbook of Linguistic Variation. Volume 3: Dutch. De Gruyter, Berlin, pp. 586-602.

Grondelaers, S., Kristiansen, T., 2013. On the need to access deep evaluations when searching for the motor of standard language change. In: Kristiansen, T., Grondelaers, S. (Eds.), Language (De)standardisation in Late Modern Europe: Experimental Studies. Novus, Oslo, pp. 9-52.

Grondelaers, S., Speelman, D., 2013. Can speaker evaluation return private attitudes towards stigmatised varieties? Evidence from emergent standardisation in Belgian Dutch. In: Kristiansen, T., Grondelaers, S. (Eds.), Language (De)standardisation in Late Modern Europe: Experimental Studies. Novus, Oslo, pp. 171-191.

Grondelaers, S., Speelman, D., Lybaert, C., Van Gent, P., 2020. Getting a (big) data-based grip on ideological change. Evidence from Belgian Dutch. Journal of Linguistic Geography 8 (1), 1-17.

Grondelaers, S., Van Gent, P., 2019. How 'deep' is Dynamism? Revisiting the evaluation of Moroccan-flavored Netherlandic Dutch. Linguistics Vanguard 5 (1), 1-11.

Grondelaers, S., Van Hout, R., 2010. Do speech evaluation scales in a speaker evaluation experiment trigger conscious and unconscious attitudes? University of Pennsylvania Working Papers in Linguistics 16 (2), 93-102.

Grondelaers, S., Van Hout, R. 2011. The standard language situation in the Low Countries: top-down and bottom-up variations on a diaglossic theme. Journal of Germanic Linguistics 23 (3), 199-243.

Grondelaers, S., Van Hout, R., 2012. Where is Dutch (really) heading? The classroom consequences of destandardization. Dutch Journal of Applied Linguistics 1 (1), 41-58.

Grondelaers, S., Van Hout, R., Van Gent, P., 2016. Destandardization is not destandardization: Revising standardness criteria in order to revisit standard language typologies in the Low Countries. Taal en Tongval 68 (2), 119-150.

Heylen, K., Van den Broeck, K., 2016. Discrimination and selection in the Belgian private rental market. Housing Studies 31 (2), 223-36.

Impe, L., Speelman, D., 2007. Vlamingen en hun (tussen)taal. Een attitudineel mixed guiseonderzoek. Handelingen van de Koninklijke Zuid-Nederlandse Maatschappij voor Taal- en Letterkunde en Geschiedenis LXI, 109-128.

Jaspers, J., 2012. Het Algemeen Nederlands: uw sociale zekerheid? Taalgebruik en taalopvattingen in processen van in- en uitsluiting. In: Kevin Absillis, K., Jaspers, J., Van Hoof, S. (Eds.), De manke usurpator: over Verkavelingsvlaams. Academia Press, Ghent, pp. 371-395.

Jaspers, J., 2013. Alhier, aldaar aan lange lansen, het Standaardnederlands. De Vlaamse tunnelvisie op emancipatie. Streven 80 (9), 784-795.

Jaspers, J., Van Hoof, S., 2013. Hyperstandardisation in Flanders: Extreme Enregisterment and Its Aftermath. Pragmatics 23 (2), 331-59.

Kang, O., Rubin, D., 2009. Reverse Linguistic Stereotyping: Measuring the Effect of Listener Expectations on Speech Evaluation. Journal of Language and Social Psychology 28 (4), 441-456.

Kristiansen, T., 2001. Two standards: one for the media and one for the school. Language Awareness 10 (1), 9-24.

Kristiansen, T., 2009. The macro-level social meaning of late modern Danish accents. Acta Linguistica Hafniensia 41, 167-192.

Kroskrity, P., 2010. Language ideologies. In: Östman, J., Verschueren, J. (Eds.), Handbook of Pragmatics Online Volume 14. John Benjamins Publishing Company, Amsterdam, pp. 1-24.

Lambert, W., Hodgson, R., Gardner, R., Fillenbaum, S., 1960. Evaluational reactions to spoken languages. The Journal of Abnormal and Social Psychology 60 (1), 44–51.

Lambert, W., Frankel, H., Tucker, G., 1966. Judging personality through speech: A French-Canadian example. Journal of Communication 16, 305-321.

Laws, E. L., Apperson, J. M., Buchert, S., Bregman, N.J., 2010. Student evaluations of instruction: When are enduring first impressions formed? North American Journal of Psychology 12 (1), 81–91.

Lybaert, C., 2017. A direct discourse-based approach to the study of language attitudes: The case of tussentaal in Flanders. Language Sciences 59, 93-116.

Lybaert, C., De Clerck, B., Plevoets, K., Loete, T., Depovere, S., 2020. De impact van taal en gender in negatieve reviews : Een experimentele studie naar gepercipieerde geloofwaardigheid en consumentengedrag. Tijdschrift Voor Taalbeheersing 42 (1), 55–94.

Marzo, S., 2016, Exploring the social meaning of contemporary urban vernaculars: Perceptions and attitudes about Citétaal in Flanders. International Journal of Bilingualism 20 (5), 501-521.

Meeus, B., 1974. A diglossic situation: Standard vs dialect. The case of Dutch in Belgium, especially in relation to urbanization. Linguistic Agency University at Trier, Trier.

Mengel, F., Sauermann, J., Zölitz, U., 2019. Gender Bias in Teaching Evaluations. Journal of the European Economic Association 17 (2), 535-566.

Myria, 2017. Migratie in cijfers en in rechten. <https://www.myria.be/files/MIGRA2017_NL_AS.pdf >.

Neuliep, J., Speten-Hansen, K., 2013. The influence of ethnocentrism on social perceptions of nonnative accents. Language & Communication 33, 167-176.

Noppe, J., Vanweddingen, M., Doyen, G., Stuyck, K., Feys, Y., Buysschaert, P., 2018. Vlaamse migratie- en integratiemonitor 2018. Agentschap Binnenlands Bestuur, Brussel. <https://www.vlaanderen.be/nl/publicaties/detail/vlaamse-migratie-en-integratiemonitor-2018>

Pantos, A.J., Perkins, A.W., 2012. Measuring Implicit and Explicit Attitudes Toward Foreign Accented Speech. Journal of Language and Social Psychology 32(1), 3-20.

Potter, J., Wetherell, M., 1987. Discourse and Social Psychology. London: Sage.

Riniolo, T., Johnson, K., Sherman, T., Misso, J., 2006. Hot or not: Do professors perceived as physically attractive receive higher student evaluations? Journal of General Psychology 133, 19-35.

Rosseel, Speelman & Geeraerts 2019a. Measuring language attitudes in context. Exploring the potential of the Personalized Implicit Association Test. Language in Society 48, 429-461.

Rosseel, L., Speelman, D., Geeraerts, D., 2019b. The relational responding task (RRT): a novel approach to measuring social meaning of language variation. Linguistics Vanguard 5 (1), 1-13.

Rubin, D., 1992. Nonlanguage Factors Affecting Undergraduates' Judgments of Nonnative English-Speaking Teaching Assistants. Research in Higher Education 33 (4), 511-531.

Rubin, D., Ainsworth, S., Cho, E., Turk, D., Winn, L., 1999. Are Greek letter social organizations a factor in undergraduates' perceptions of international instructors? International Journal of Intercultural Relations 23 (1), 1–12.

Rubin, D., Smith, K., 1990. Effects of accent, ethnicity, and lecture topic on undergraduates' perceptions of non-native English speaking teaching assistants. International Journal of Intercultural Relations 14, 337-353.

Ryan, E., Carranza, M., 1975. Evaluative reactions of adolescents toward speakers of standard English and Mexican American accented English. Journal of Personality and Social Psychology 31 (5), 855-863.

Silverstein, M., 1979. Language Structure and Linguistic Ideology. In: Clyne, P., Hanks, W., Hofbauer, C. (Eds.), The Elements. Chicago Linguistics Society, Chicago, pp. 193-248.

Silverstein, M., 2003. Indexical order and the dialectics of sociolinguistic life. Language and Communication 23, 193-229.

Soukup, B., 2013. The measurement of 'language attitudes' – A reappraisal from a constructionist perspective. In: Kristiansen, T., Grondelaers, G. (Eds), Language (de)standardisation in late modern Europe: Experimental studies. Novus, Oslo, pp. 251-266.

Spooren, P., Brockx, B., Mortelmans, D., 2013. On the validity of student evaluation of teaching: The state of the art. Review of Educational Research 83 (4), 598–642.

Subtileru, N., 2015. "She does have an accent but…": Race and language ideology in students' evaluations of mathematics instructors on RateMyProfessors.com. Language in Society 44, 35-62.

Van der Bracht, K., Coenen, A., Van de Putte, B., 2015. The Not-in-My-Property Syndrome: The Occurrence of Ethnic Discrimination in the Rental Housing Market in Belgium. Journal of Ethnic and Migration Studies 41 (1), 158-75.

Van de Velde, H., Houtermans, M., 1999. Vlamingen en Nederlanders over de uitspraak van nieuwslezers. In: Huls, E., Weltens, B. (red.), Artikelen van de Derde Sociolinguïstische Conferentie. Eburon, Delft, 451-462.

Vandekerckhove, R., 2000. Structurele en sociale aspecten van dialectverandering. De dynamiek van het Deerlijkse dialect. Koninklijke Academie voor Nederlandse Taal- en Letterkunde, Ghent.

Vandekerckhove, R., 2009. Dialect loss and dialect vitality in Flanders. International Journal of the Sociology of Language 196/197, 73-97.

Vandekerckhove, R., Cuvelier, P., 2007. The perception of exclusion and proximity through the use of standard Dutch, 'tussentaal' and dialect in Flanders. In: Cuvelier, P., Du Plessis, T., Meeuwis, M., Teck, L. (Eds..), Multilingualism and exclusion. Policy, practice and prospects. Van Schaik Publishers, Pretoria, pp. 241-256.

Van Hoof, S., 2018. Civilization versus commerce: on the sociolinguistic effects of the deregulation of the TV market on Flemish public service broadcasting. Language policy 17(2), 199-216.

Van Hoof, S., Nyssen, S., Kanobana, S., 2020. 'If they could, they would put them on a drip with Dutch': language learning and the professional integration of migrants in Flanders. International Journal of the Sociology Of Language 264, 73-94.

Verhaeghe, P.P., Praktijktesten Nu, 2018. Etnische discriminatie op de private huurwoningmarkt in Antwerpen. Free University of Brussels, Brussels.

Weichselbaumer D., 2020. Multiple Discrimination against Female Immigrants Wearing Headscarves. Industrial and Labor Relations Review 73 (3), 600-627.

Woolard, K., Schieffelin, B., 1994. Language Ideology. Annual Review of Anthropology 23, 55-82.

Yi, H., Phelps, J., Smiljanic, R., Chandrasekaran, B., 2013. Reduced efficiency of audiovisual integration for nonnative speech. The Journal of the Acoustical Society of America 134 (5), 387-393.

Yi, H. G., Smiljanic, R., Chandrasekaran, B., 2014. The neural processing of foreign-accented speech and its relationship to listener bias. Frontiers in human neuroscience 8, 768.

Zahn, C., Hopper, R., 1985. Measuring language attitudes: The Speech Evaluation Instrument. Journal of Language and Social Psychology 4 (2), 113-123.

Zheng, Y., Samuel, A., 2017. Does seeing an Asian face make speech sound more accented? Attention, Perception, & Psychophysics 79 (6), 1841-59.

Appendices

Appendix 1: Transcripts of speech clips

Belgian Standard Dutch

De Frans-Spaanse oorlog die zal bijna een kwart eeuw duren en die kent uiteindelijk een heel interessante uitkomst, die zomaar uit de mouw kwam geschud van Lodewijk de veertiende. En je moet mij nu eventjes excuseren, ik ga een beetje in detail gaan, maar 't is revelerend om de mechanismen bloot te leggen. Waar zal Lodewijk de veertiende de koning van Frankrijk op aansturen? Spanje was zeer duidelijk aan de verliezende hand, en Spanje wou eigenlijk ook vrede na zo veel jaren oorlog. Je kan je al wel voorstellen wat Lodewijk de veertiende zal willen van Spanje. Hij zal natuurlijk voorstellen om een paar gebieden over te dragen aan Frankrijk. Maar hij gaat ook aansturen op een huwelijk. Op wie heeft Lodewijk de veertiende zijn oog laten vallen? Dat is… op… 'et enige overblijvende kind uit 'et eerste huwelijk van Filips de vierde van Spanje. Dat was zijn tegenhanger in Spanje. Filips de vierde. Uit die zijn huwelijk met Isabella van Bourbon zijn er twee kinderen voortgekomen. Het eerste is de troonsopvolger, Balthasar Carlos. In 1629 wordt die geboren… maar… hij overlijdt in 1646. Het enige kind uit dat huwelijk dat overblijft dat is een dochter: Maria Theresia. En het is op haar dat Lodewijk de veertiende zijn oog zal laten vallen. Spanje gaat dat weigeren. Waarom gaat Spanje dat weigeren? Omdat dat zou betekenen dat de enige nog levende erfgenaam… trouwt met de koning van Frankrijk. En dat hun nageslacht zou regeren over Spanje én Frankrijk. Dat zou dus betekenen dat Filips de vierde eigenlijk Spanje op termijn schenkt aan Lodewijk de

veertiende. Dat kan natuurlijk niet voor Spanje, zo'n scenario dat is de grootste schrik van Madrid. Dus… In Madrid zegt men: geen sprake van. Maar Filips de vierde trouwt in 1649 een tweede keer, met aartshertogin Maria Anna van Oostenrijk. En daar komt in 1657 een zoon uit. Een troonsopvolger: Felipe Prospero. Op dat moment is Madrid wél bereid om te gaan spreken met Parijs, omdat men ervan uitgaat dat met Felipe Prospero, de zoon uit 'et tweede huwelijk, de erfopvolging gegarandeerd is. Dus vanaf 1657 wil Madrid wel praten over een huwelijk in het kader van de vrede. En dat is wat dan zal gebeuren. In 1659 wordt er vrede gesloten en één van de bepalingen van de vrede is dat Lodewijk de veertiende kan trouwen met de dochter van Filips de vierde, het enige overblijvende kind uit 'et eerste huwelijk. Voilà, een koppel. En dat is het einde van de Frans-Spaanse oorlog.

Colloquial Belgian Dutch

De Frans-Spaansen oorlog die ga bijna een kwart eeuw duren en die kend uiteindelijk een 'eel interessante uitkomst, die zomaar uit de mouw kwam geschud van Lodewijk de veertiende. En ge moet mij nu efkes excuseren, ik ga een beetjen in detail gaan, maar 't is revelerend om de mechanismen bloot te leggen. Waar ga Lodewijk de veertiende de koning van Frankrijk op aansturen? Spanje was eel duidelijk aan de verliezende hand, en Spanje wou eigenlijk ook vrede na zo veel jaren oorlog. Ge kund u al wel voorstellen wat da Lodewijk de veertiende ga willen van Spanje. 'ij ga natuurlijk voorstellen van een paar gebieden over te dragen aan Frankrijk. Maar 'ij gaad ook aansturen op een huwelijk. Op wie 'eeft Lodewijk de veertiende zijn oog laten vallen? Dad is… op… 'et enigste overblijvende kind uit 'et eerste huwelijk van Filips de vierde van Spanje. Da was zijnen tegenhanger in Spanje. Filips de vierde. Uit dieje zijn huwelijk me Isabella van Bourbon zijn d'r twee kinderen voortgekomen. Het eerste is den troonsopvolger, Balthasar Carlos. In 1629 wordt dieje geboren… maar… diejen overlijdt in 1646. 'et enigste kind uit dad 'uwelijk da overblijft dad is een dochter: Maria Theresia. En het is op haar da Lodewijk de veertiende zijn oog ga laten vallen. Spanje gaat da weigeren. Waarom ga Spanje da weigeren? Omdat da zou betekenen dat den enigste nog levenden erfgenaam… trouwt met de koning van Frankrijk. En da hun nageslacht zou regeren over Spanje én Frankrijk. Da zou dus betekenen da Filips de vierde eigenlijk Spanje op termijn schenkt aan Lodewijk de veertiende. Ja da kan natuurlijk nie voor Spanje, zo'n scenario dad is de grootste schrik van Madrid. Dus… In Madrid zegt men: geen sprake van. Maar Filips de vierde trouwt in 1649 nen tweede keer, me aartshertogin Maria Anna van Oostenrijk. En daar komt in 1657 ne zoon uit. Nen troonsopvolger: Felipe Prospero. Op da moment is Madrid wél bereid om te gaan spreken me Parijs, omda men d'rvan uitga da me Felipe Prospero, de zoon uit da tweede huwelijk, de erfopvolging gegarandeerd is. Dus vanaf 1657 wilt Madrid wel praten over een huwelijk in het kader van de vrede. En dad is wat dan ook ga gebeuren. In 1659 wordt er vrede gesloten en één van de bepalingen van de vrede is da Lodewijk de veertiende kan trouwen met de dochter van Filips de vierde, het enigste overblijvende kind uit 'et eerste huwelijk. Voilà, een koppel. En da is 'et einde van de Frans-Spaansen oorlog.

Appendix 2: Standard vs. non-standard features in speech clips

Variable	**Standard variant**	**Non-standard variant**
2nd person singular pronoun	*je*	*ge*
diminutive	*-je* (*eventjes*)	*-ke* (*efkes*)
absence vs. presence of expletive 'dat' after question words in subordinate interrogative sentence	Ø (*wat*)	'dat' (*wat da*)

word-final (t) in function words	[t] (*wat, dat, met, omdat, niet*)	deletion or voicing (*wa, da, me, omda, nie, dad*)
word-initial (h)	[h]	deletion (*eel, eeft, et, uwelijk, ij*)
inflection of adjective *enig*	-e *(enige)*	-ste *(enigste)*
conjugation of 3rd person singular present indicative verb forms	Ø (*wil*), -t *(gaat*)	*-t* (*wilt*), Ø (*ga*)
adnominal inflection before male nouns	Ø or -e (*zijn tegenhanger, een zoon, de troonsopvolger, de levende erfgenaam, een tweede keer, Frans-Spaanse oorlog*)	-en (*zijnen tegenhanger, ne zoon, den troonsopvolger, den levenden erfgenaam, nen tweede keer, Frans-Spaansen oorlog*)
distal masculine demonstrative pronoun	Ø (uninflected *die*)	-e(n) (inflected *dieje, diejen*)
absence vs. insertion of a linking-*n* between word-final and word-initial vowels	Ø (*beetje*)	-n (*beetjen*)
adverb *er*	*er, ervan*	*d'r, d'rvan*

Appendix 3: Scales

Likert statements (in the order they were presented in in the survey)

Original Dutch statements	**English translation**
Ik denk dat deze lesgever veel professionele ervaring heeft.	I think this instructor has a lot of professional experience.
Deze lesgever komt zelfverzekerd over.	This instructor comes across as self-confident.
Ik vind deze lesgever geloofwaardig.	I think this instructor is credible.
Ik vind deze lesgever sympathiek.	I think this instructor is nice.
Ik vind deze lesgever cool.	I find this instructor cool.
Ik vind deze lesgever goed verstaanbaar.	I find this instructor easy to understand.
Ik denk dat deze lesgever haar vakgebied kent.	I think this instructor knows her field.
Ik denk dat deze lesgever grappig uit de hoek kan komen	I think this instructor can be funny.
Ik denk dat deze lesgever gezag heeft.	I think this instructor has authority.
Ik denk dat deze lesgever betrouwbaar is.	I think this instructor is reliable.
Ik denk dat deze lesgever behulpzaam is.	I think this instructor is helpful.
Ik denk dat deze lesgever intelligent is.	I think this instructor is intelligent.
Ik vind het aangenaam om van deze persoon les te krijgen.	I enjoy being taught by this person.
Ik denk dat deze lesgever een succesvolle academische carrière heeft.	I think this instructor has a successful academic career.
Ik denk dat deze lesgever assertief is.	I think this instructor is assertive.
Ik heb de uitleg van deze lesgever goed begrepen.	I have understood the explanation of this instructor well.
Ik vind deze lesgever overtuigend.	I find this instructor convincing.
Ik denk dat deze lesgever een warme persoonlijkheid heeft.	I think this instructor has a warm personality
Ik vind deze lesgever entertainend.	I find this instructor entertaining.

Ik vind dat deze lesgever de materie helder uitlegt.	I think this instructor explains the matter clearly.
Ik denk dat studenten opkijken naar deze lesgever.	I think students look up to this instructor.
Ik zou graag van deze persoon les krijgen.	I would like to be taught by this person.
Ik denk dat deze lesgever populair is bij studenten.	I think this instructor is popular with students.
Ik vind dat deze lesgever duidelijk spreekt.	I think this instructor speaks clearly.

Semantic differential items (in the order they were presented in in the survey)

Original Dutch statements	**English translation**
Deze lesgever spreekt niet mooi vs. Deze lesgever spreekt mooi.	This instructor does not speak beautifully vs. This instructor speaks beautifully.
Deze lesgever heeft geen buitenlands accent vs. Deze lesgever heeft een buitenlands accent.	The instructor does not have a foreign accent vs. The instructor has a foreign accent.
Ik vind deze lesgever niet aantrekkelijk vs. Ik vind deze lesgever aantrekkelijk.	I don't find this instructor attractive vs. I find this instructor attractive.
Deze lesgever spreekt geen goed Nederlands vs. Deze lesgever spreekt goed Nederlands.	This instructor speaks sloppy vs. This instructor speaks properly.
Deze lesgever is geen moedertaalspreker van het Nederlands vs. Deze lesgever is een moedertaalspreker van het Nederlands.	This instructor is not a native speaker of Dutch vs. This instructor is a native speaker of Dutch
Ik vind deze lesgever niet knap vs. Ik vind deze lesgever knap.	I don't think this instructor is pretty vs. I think this instructor is pretty.
Deze lesgever spreekt onverzorgd vs. Deze lesgever spreekt verzorgd.	This teacher doesn’t speak proper Dutch vs. This teacher speaks proper Dutch
Ik vind dat deze lesgever er niet goed uitziet vs. Ik vind dat deze lesgever er goed uitziet.	I think this instructor isn’t good-looking vs. I think this instructor is good-looking.
Deze lesgever spreekt niet met een regionaal accent vs. Deze lesgever spreekt met een regionaal accent.	The instructor does not have a regional accent vs. The instructor has a regional accent.
Deze lesgever spreekt dialect vs. Deze lesgever spreekt Algemeen Nederlands	This instructor speaks a dialect vs. This instructor speaks Standard Dutch.
Ik vind deze lesgever niet mooi vs. Ik vind deze lesgever mooi.	I don’t think this instructor is beautiful vs. I think this instructor is beautiful.
Deze lesgever spreekt geen correct Nederlands vs. Deze lesgever spreekt correct Nederlands.	This instructor does not speak correct Dutch vs. This instructor speaks correct Dutch.

Appendix 4: Estimated Marginal Means Comparison

Dependent Variable	**Cond. i**	**Cond. j**	**Mean Diff (i-j)**	**se**	***p***	**95% Conf. Int.**	
Professionalism	1	3	-.370*	0.089	.000	-0.545	-0.195
	2	3	-.275*	0.083	.001	-0.439	-0.112
	3	1	.370*	0.089	.000	0.195	0.545
		2	.275*	0.083	.001	0.112	0.439
		4	.302*	0.098	.002	0.108	0.495

		5	.308*	0.099	.002	0.113	0.503
		6	.255*	0.091	.005	0.077	0.433
	4	3	-.302*	0.098	.002	-0.495	-0.108
	5	3	-.308*	0.099	.002	-0.503	-0.113
	6	3	-.255*	0.091	.005	-0.433	-0.077
Comprehensibility	1	2	-.370*	0.109	.001	-0.585	-0.156
		3	-.798*	0.127	.000	-1.048	-0.548
		4	-.366*	0.133	.006	-0.627	-0.104
		5	-.682*	0.134	.000	-0.946	-0.418
		6	-.310*	0.121	.011	-0.548	-0.072
	2	1	.370*	0.109	.001	0.156	0.585
		3	-.428*	0.119	.000	-0.661	-0.194
		5	-.312*	0.126	.014	-0.56	-0.064
	3	1	.798*	0.127	.000	0.548	1.048
		2	.428*	0.119	.000	0.194	0.661
		4	.432*	0.14	.002	0.156	0.709
		6	.488*	0.129	.000	0.233	0.743
	4	1	.366*	0.133	.006	0.104	0.627
		3	-.432*	0.14	.002	-0.709	-0.156
		5	-.316*	0.147	.032	-0.605	-0.028
	5	1	.682*	0.134	.000	0.418	0.946
		2	.312*	0.126	.014	0.064	0.56
		4	.316*	0.147	.032	0.028	0.605
		6	.372*	0.136	.007	0.104	0.64
	6	1	.310*	0.121	.011	0.072	0.548
		3	-.488*	0.129	.000	-0.743	-0.233
		5	-.372*	0.136	.007	-0.64	-0.104
Physical attractiveness	1	2	.347*	0.157	.028	0.038	0.657
		3	.878*	0.184	.000	0.517	1.239
		4	1.100*	0.192	.000	0.723	1.477
		6	.557*	0.175	.002	0.213	0.901
	2	1	-.347*	0.157	.028	-0.657	-0.038
		3	.530*	0.171	.002	0.193	0.867
		4	.753*	0.18	.000	0.399	1.107
	3	1	-.878*	0.184	.000	-1.239	-0.517
		2	-.530*	0.171	.002	-0.867	-0.193
		5	-.565*	0.204	.006	-0.967	-0.163
	4	1	-1.100*	0.192	.000	-1.477	-0.723
		2	-.753*	0.18	.000	-1.107	-0.399
		5	-.787*	0.212	.000	-1.204	-0.371
		6	-.543*	0.195	.006	-0.926	-0.16

	5	3	.565*	0.204	.006	0.163	0.967
		4	.787*	0.212	.000	0.371	1.204
	6	1	-.557*	0.175	.002	-0.901	-0.213
		4	.543*	0.195	.006	0.16	0.926
Standardness	1	2	.609*	0.124	.000	0.365	0.852
		4	.824*	0.151	.000	0.528	1.121
		6	.579*	0.138	.000	0.309	0.85
	2	1	-.609*	0.124	.000	-0.852	-0.365
		3	-.840*	0.135	.000	-1.106	-0.575
		5	-.581*	0.143	.000	-0.863	-0.3
	3	2	.840*	0.135	.000	0.575	1.106
		4	1.056*	0.159	.000	0.742	1.37
		6	.811*	0.147	.000	0.522	1.1
	4	1	-.824*	0.151	.000	-1.121	-0.528
		3	-1.056*	0.159	.000	-1.37	-0.742
		5	-.797*	0.166	.000	-1.125	-0.47
	5	2	.581*	0.143	.000	0.3	0.863
		4	.797*	0.166	.000	0.47	1.125
		6	.552*	0.155	.000	0.248	0.856
	6	1	-.579*	0.138	.000	-0.85	-0.309
		3	-.811*	0.147	.000	-1.1	-0.522
		5	-.552*	0.155	.000	-0.856	-0.248

Note. **Confidence interval Bonferroni adjusted for multiple comparisons**