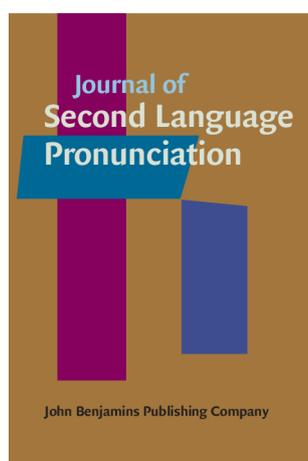


Disentangling Professional Competence and Foreign Accent

Cesar Teló,¹ Pavel Trofimovich,² Mary Grantham O'Brien^{1*}

¹University of Calgary | ²Concordia University



Teló, C., Trofimovich, P., & O'Brien, M. (2022). Disentangling professional competence and foreign accent. *Journal of Second Language Pronunciation*. Published online 29 September 2022. <https://doi.org/10.1075/jslp.22013.tel>

Abstract

This study examined listeners' evaluations of first (L1) and second language (L2) English speech in work-related contexts. Ninety-six English-speaking listeners from Calgary rated audio recordings of 12 English speakers (6 L1 English, 6 L1 Tagalog) along three continua capturing one professional (competence), one experiential (treatment preference), and one linguistic (comprehensibility) dimension. The audio recordings additionally differed in terms of job prestige (high vs. low) and performance level (high vs. low). Compared to English speakers, Tagalog speakers were rated as less competent and comprehensible overall, and listeners wished to be treated more like the clients in scenarios recorded by English than Tagalog speakers, with all effects magnified for speakers with heavier foreign accents. Nonetheless, listeners generally evaluated English and Tagalog speakers similarly in low-prestige and in low-performance scenarios, but rated low performance less negatively in low-prestige positions. Findings demonstrate highly nuanced accent bias in work-related contexts.

Keywords: accentism, foreign accent, speech assessment, workplace, performance, prestige, second language, competence, treatment preference, comprehensibility

1. Introduction

Imagine visiting your family doctor. At the appointment, the doctor requests lab tests and then calls you a week later to tell you about the results. The doctor sounds rushed, uses technical jargon, and suggests that you go online for additional information. Your reactions to this doctor would almost certainly be less than favorable, as you might experience difficulty understanding the message, question the doctor's competence, or feel unhappy to be treated that way. How would this doctor compare to another who takes the time to interpret and explain the results and provides you with clear instructions for next steps? More importantly,

what if one or both these professionals spoke with a foreign accent, which would mark them as a second language (L2) speaker? Finally, would it make a difference if, instead of a doctor, a server at a restaurant performed their job more or less competently? These scenarios illustrate our main goal in this study, which was to examine whether the presence and the degree of a foreign accent affects how listeners evaluate a speaker's professional performance. We addressed this goal by having listeners evaluate speakers in simulated job scenarios, where native and L2 speakers of English demonstrated either high or low professional performance for jobs that varied in prestige, such as a doctor versus a server.

2. Background Literature

It is common knowledge that listeners make inferences about speakers based on their accent, which broadly refers to segmental and suprasegmental characteristics contributing to listener perceptions of how closely a given speech pattern matches the expected variety (Derwing & Munro, 2015). These inferences are more often negative than positive, and they tend to arise readily in response to foreign accents, which mark the speaker as a member of a minority language group and contribute to the perception of “otherness” (Dragojevic, Gasiorek, & Giles, 2016; Giles & Watson, 2013). For example, listeners who belong to the majority language group often ascribe negative judgments to L2 speakers, evaluating those with heavy accents less favorably in terms of their status traits, such as knowledge, competence, and intelligence, and their personal characteristics, such as friendliness, honesty, and assertiveness (Baquiran & Nicoladis, 2020; Cargile & Giles, 1998; Hosoda, Stone-Romero, & Walter, 2007; Lindemann, 2003; Nelson Jr., Signorella, & Botti, 2016). While some accent-driven judgments are harmless in that they do not carry negative consequences, many have been associated with stereotyping, stigma, and outright discrimination of L2 speakers (Lippi-Green, 2012; Moyer, 2013).

Accentism—a particular type of linguistic bias in which inequitable treatment is based on a speaker’s accent—may be particularly pronounced in business and professional settings. In pre-employment situations, for example, job applicants from minority groups are approximately 50% less likely to be invited to a job interview (Zschirnt & Ruedin, 2016). When L2 speakers get interviewed, their pronunciation is again a source of bias (Deprez-Sims & Morris, 2010; Segrest Purkiss, Perrewé, Gillespie, Mayes, & Ferris, 2006), such that speaking with a foreign accent results in classification, judgment and, ultimately, punishment of job applicants (Roberts, 2021). Even trained professionals, such as human resource (HR) personnel, are prone to accentism. For instance, HR specialists tended to downgrade the overall job suitability of foreign-accented applicants (Almeida, Fernando, Hannif, & Dharmage, 2015), and workers with managerial experience considered foreign-accented interviewees less adequate for customer-facing jobs than for non-customer-facing positions (Timming, 2017).

Accents also seem to play a role in workplace practices, with many L2 speakers earning lower wages than native speakers (Dávila, Bohara, & Saenz, 1993; Hamilton, Goldsmith, & Darity, 2008; Reitz & Sklar, 1997). Foreign accents have been described as distracting to business communication (Mai & Hoffmann, 2014) and an impediment to job-related tasks (Ramjattan, 2019). In Canada, which is the context of our study, Chinese-accented physicians are judged as less competent than physicians speaking with standard Canadian accents (Baquiran & Nicoladis, 2020), and foreign-accented nurses are ostracized more often for their pronunciation than foreign-accented doctors (Neiterman & Bourgeault, 2015), which implies that accent bias depends on job prestige. In the United States, Black African-born nurses report that their accents often lead patients to see them as lacking adequate professional skills, preparedness, and intelligence, and trigger belittling behaviors from coworkers (Iheduru-Anderson, 2020; see also Ramjattan, 2019).

However, not all L2 speakers are judged similarly based on their accents. For instance, in an early study (Kalin, Rayko, & Love, 1980), German speakers of English were preferred over West Indian and South Asian English speakers for high-status jobs. Similarly, when Spanish-accented speakers were presented to listeners as belonging to a lower social class, these speakers were particularly downgraded in listener evaluations compared to native English speakers and Spanish-accented speakers believed to belong to higher social standing (Ryan & Sebastian, 1980). Perception of lower social class was also associated for listeners with a speaker being less successful, for example, in terms of gaining a job promotion (Stewart, Ryan, & Giles, 1985). More recently, Dragojevic and Goatley-Soan (2022) showed a similar pattern of findings in relation to other listener-assessed attributes, where American listeners assigned higher ratings of status (e.g., competence, intelligence) and personality (e.g., friendliness, warmth) to L2 English speakers from Western Europe (e.g., France, Germany) but downgraded speakers from South and East Asia (e.g., China, India, Vietnam), effectively categorizing L2 speakers into preferred (Western European) versus non-preferred (Latin, Middle Eastern, South and East Asian) groups. This pattern of speaker categorization was closely matched by listener ratings of the speakers' comprehensibility, which is a scalar measure of how difficult it is for listeners to understand a speaker (Derwing & Munro, 2015). The speakers from preferred groups were rated as easier to understand than those from non-preferred groups, and within each group, the speakers' comprehensibility and their status and personality ratings were correlated, with coefficients as high as .93. Thus, difficulty understanding a foreign-accented speaker can trigger low ratings of competence, and vice versa, low ratings of competence can aggravate potential comprehensibility issues.

3. The Current Study

Prior work on accentism has revealed that L2 speakers are generally disadvantaged in various pre- and in-employment contexts and that some speaker groups can be more

vulnerable to bias than others. However, it is presently unclear if accent bias is undifferentiated, in the sense that L2 speakers are downgraded in their evaluations, relative to native speakers, regardless of their demonstrated job-relevant competence or performance level. Furthermore, accent bias might also depend on job status (Neiterman & Bourgeault, 2015), considering that L2 speakers are preferred for low- over high-prestige jobs (Brennan & Brennan, 1981; Kalin & Rayko, 1978; Kalin et al., 1980) and for semi-skilled over supervisory roles (De La Zerda & Hopper, 1979; Iheduru-Anderson, 2020). Finally, although prior work on accentism has centered on perceived competence and employability of L2 speakers (Baquiran & Nicoladis, 2020; Timming, 2017), studies tapping into a broader range of workplace-related attributes are still scarce. Besides perceived competence, it would also be important to consider other performance indicators, including those that capture listener experience or satisfaction (Darian, Tucci, & Wiman, 2001; Walsh, Gouthier, Gremler, & Brach, 2012) and those that describe L2 speakers' language, such as comprehensibility (Dragojevic & Goatley-Soan, 2022).

Thus, in this study, we extended previous work on accentism in work-relevant contexts by examining the performance of native and L2 speakers of English in simulated job scenarios which demonstrated high or low job-specific performance levels for occupations that were high or low in prestige. The scenarios were rated by listeners for three measures that captured professional (competence), experiential (degree to which listeners would want to be treated as depicted in the scenario), and linguistic (comprehensibility) dimensions of each performance. The study was conducted in Calgary (Alberta) and targeted Filipino speakers of L2 English. As a group, Filipinos tend to immigrate in search of employment and better economic conditions (Go, 1998). Although Filipinos are valued for their expertise and work ethic (Coloma, McElhinny, Tungohan, Catungal, & Davidson, 2012; Lightman, Banerjee, Tungohan, de Leon, & Kelly, 2021), they often occupy low-status, low-paying

(mostly manual, service-related) jobs unwanted by the locals (Semyonov & Gorodzeisky, 2004), and like many L2 speakers, seem to experience workplace discrimination (de Castro, Gee, & Takeuchi, 2008). In Calgary, Filipinos make up nearly 6% of the population, with Tagalog as the most widely spoken mother tongue other than English (Statistics Canada, 2017), which means that Calgarians frequently interact with Tagalog-accented speakers when acquiring goods and services.

Our study was guided by the following question: Are listeners' professional (competence), experiential (treatment preference), and linguistic (comprehensibility) reactions to speakers in job scenarios associated with speakers' linguistic background (English vs. Tagalog), job prestige (high vs. low), and performance level (high vs. low)? On the basis of previous research, we predicted that L2 speakers' competence ratings would be downgraded overall (Iheduru-Anderson, 2020), with listeners showing a preference for native speakers (Moyer, 2013). Although listeners from many backgrounds tend to understand Tagalog-accented English quite well (Dayag, 2007; Li & Chen, 2019), we nonetheless expected that L2 speakers would elicit lower comprehensibility ratings, compared to native speakers, especially because Filipinos, as South-East Asian speakers of L2 English, might be categorized as a non-preferred group (Dragojevic & Goatley-Soan, 2022). We also expected that all speakers, including L2 speakers, in relatively low-prestige positions would be assigned less favorable ratings than those in high-prestige jobs (Neiterman & Bourgeault, 2015; Ryan & Sebastian, 1980; Stewart et al., 1985). While we know of no previous research that has investigated the extent to which native and L2 speakers are assigned differing penalties for poor job performance, we predicted that listeners would show a general preference for speakers demonstrating a high level of job-specific performance.

4. Method

4.1. Recordings

The target audios were recorded by six native English and six native Tagalog speakers (three females and three males per group), all residents of Calgary. The recordings were elicited through 12 scripted scenarios, where each included a brief job-related comment narrated by a hypothetical professional employed in that job. Half of the scenarios involved low-prestige jobs (cleaner, salesperson, server) while the other half involved high-prestige jobs (doctor, lawyer, professor), with job status determined through occupational prestige scores and socioeconomic indexes (Goyder, Guppy, & Thompson, 2003; Hauser & Warren, 1997). Within each occupation, one script exemplified high-level performance while the other illustrated low-level performance by each professional. The scenarios portrayed common situations faced by Calgarians in their daily lives when interacting with the selected low- and high-prestige job professionals. For instance, a high performance doctor script featured a phone message to a patient in which the doctor explained the results of the patient's lab tests, providing detailed yet layperson-friendly explanations and informing the patient that a prescription was ready because the lab results indicated a possible bladder infection. In contrast, a low performance doctor script provided similar information, except that the doctor sounded rushed, reporting the results using medical jargon and suggesting that the patient contact a nurse for a prescription because the doctor was too busy to write it.¹

Before recording the audios, to ensure that the scripts illustrated different performance levels, they were evaluated by 53 raters from the introductory linguistics participant pool who assessed their general quality and provided brief comments about them. Following this initial feedback, the scripts were revised and evaluated again by 18 additional raters from the same participant pool. The raters read each script and used 1–9 scales (1 = *not at all*, 9 = *very much*) to assess the person in it for pleasantness, job effectiveness, and

sensitivity to the needs of others and finally to estimate their familiarity with each scenario. Compared to low-performance scripts, high-performance scripts illustrated a professional who was more pleasant ($M = 2.95$ vs. 7.82), $t(16) > 3.31$, $p < .004$, more effective ($M = 2.95$ vs. 8.18), $t(16) > 3.30$, $p < .005$, and more sensitive to the needs of others ($M = 1.85$ vs. 7.70), $t(16) > 2.93$, $p < .01$. However, both high- and low-performance scripts were rated as similarly familiar ($M = 5.07$ vs. 6.34), $t(16) < 1.97$, $p > .07$. Thus, the high- versus low-performance scenarios were distinguishable in performance level yet comparable in rater-assessed familiarity.

The final version of the scripts (see Appendix A), which ranged between 95 and 119 words ($M = 109.92$), were recorded by speakers using *Praat* (Boersma & Weenink, 2021) in individual online meetings with a researcher. Prior to the meeting, the speakers were sent the scripts by e-mail and instructed to read and familiarize themselves with the passages. They were directed to read each script as naturally as possible and to speak at a normal pace as if they were leaving a voicemail message, with the opportunity to record each passage multiple times. The recording that featured the most optimal pace (not too fast, not too slow), that did not include multiple, obtrusive false starts, self-corrections, and repetitions, and that sounded natural, as judged by the researcher, was considered for the final set of recorded scenarios.

Finally, to ensure that the recordings by native Tagalog and English speakers were comparable within but different between the two sets, we compared the ratings of accentedness assigned to all recorded scripts by the 96 listeners (see below) as part of the experimental procedure. The six Tagalog speakers' accentedness ratings ranged between 32.76 and 60.39 ($M = 49.61$) on a 100-point scale (where 100 = *not accented at all*), whereas the six English speakers' accentedness ratings ranged between 93.93 and 97.55 ($M = 95.86$), with a reliable difference between the two sets, $t = 43.26$, $p < .001$, $M_{diff} = 46.08$, 95% CI [43.99, 48.17], $d = 2.54$. Therefore, insofar as accentedness ratings capture variations in a

speaker's pronunciation (at the levels of segments and prosody) and voice quality (Derwing & Munro, 2015), the recordings by Tagalog and English speakers appeared to represent distinct pronunciation performances.

4.2. Listeners

To evaluate the audios, 96 listeners ($M_{age} = 38.76$ years, $SD = 15.88$) were recruited through advertisements distributed through the university's official participant recruitment page, Facebook groups, and word of mouth, using age (18 years old or older) and current city of residence (Calgary) as inclusion criteria. Listener recruitment specifically targeted community members rather than university students, in keeping with this study's goal to examine how Calgarians evaluate professional job performances by native English and Tagalog speakers. Listeners' first language (L1) backgrounds included English (73%), Cantonese, Mandarin, Vietnamese, Korean, French, Spanish, Ukrainian, Dutch, and German (26%). One listener reported L1 Tagalog while another did not state their L1. Listeners reported using English frequently ($M = 94.03$, $SD = 10.83$) in their daily lives (0 = *not at all*, 100 = *all the time*). They also self-rated their familiarity with accented English ($M = 7.65$, $SD = 1.76$) as fairly high (1 = *not familiar at all*, 9 = *very familiar*). All but one reported having normal hearing. Excluding that listener's data resulted in no change in the findings; therefore, the entire listener sample was considered for analysis.

Although a substantial number of listeners included L1 English speakers, our a priori decision was to obtain a realistic picture of how Calgarians perceive native and L2 speakers in job-related scenarios, so no separate analyses for L1 and L2 English-speaking listeners were performed. First, according to recent meta-analytic evidence, native and L2-speaking listeners appear highly comparable in the speech ratings they provide for linguistic dimensions such as accentedness and comprehensibility, meaning that listeners demonstrate high rating consistency regardless of whether they are trained raters (i.e., with prior

coursework in linguistics and phonetics and teaching experience), whether they are naïve, layperson listeners, or whether they are L2 speakers themselves (Saito, 2021). And secondly, from a practical perspective, it was challenging to make the distinction between a listener's native versus non-native status in a multilingual and multicultural context such as Calgary, where residents are exposed to a great deal of linguistic diversity because the city of 1.3 million is home to a sizeable proportion of immigrants (30%) representing over 240 ethnic communities (Statistics Canada, 2017). In fact, as a group, the 96 listeners nearly perfectly illustrated the linguistic landscape of Calgary, with 26% of listeners reporting a mother tongue other than English. Nevertheless, to account for potential impact of listeners' self-reported knowledge of more than one language (as described below), we included in all analyses a control covariate capturing the number of languages each listener reported speaking (ranging between 1 and 5 in our sample).

4.3. *Materials and Procedure*

Listeners evaluated audio-recorded scripts remotely, using a *Qualtrics* web-based interface. The interface included an audio recording followed by several 100-point sliding scales which contained only endpoint descriptors. Three scales captured listeners' professional, experiential, and linguistic reactions to the speaker in each scenario. The professional scale targeted the speaker's competence (This person seems competent... *not at all–very much*). The experiential scale captured listeners' perception of the degree to which they themselves wanted to be treated as described in the script (I would want to be treated as described in this script... *not at all–very much*). The linguistic scale focused on the speaker's comprehensibility (This person is... *hard to understand–easy to understand*). An additional scale targeted the speaker's accentedness, which was included as a predictor in all subsequent analyses to capture between-speaker variation in pronunciation (This person is... *heavily accented–not accented at all*). As part of a larger project, other scales focused on the

speaker's personal traits (e.g., pleasantness) and interpersonal skills (e.g., politeness); these data, which targeted a conceptually different question and whose treatment would not be possible due to space limitations, are not discussed further. Comprehensibility was defined for listeners as the degree of effort it takes to understand what the speaker is saying, whereas accentedness was defined as the degree to which the speaker's pronunciation departs from the expected native-speaker variety (Derwing & Munro, 2015). The other two ratings, considered to be intuitive to a layperson listener, were not introduced through a definition but were instead preceded by a contextualizing sentence (i.e., This person seems competent).

The final set of recorded scripts contained 48 files, where each of the 12 speakers was randomly chosen to contribute audios for one high-prestige job (low- and high-level performance) and one low-prestige job (low- and high-level performance), for a total of four audios per speaker. Because the scripts were recorded by speakers who varied in their gender and L1, and because the scripts illustrated both high- and low-performance scenarios for one of the six jobs, it was ensured that no listener evaluated high- and low-performance scenarios for the same job, heard the same speaker's voice twice, or heard an unbalanced number of L1 versus L2 speaker recordings. Therefore, the 48 audios (12 speakers \times 4 scripts each) were organized in eight balanced experimental lists containing six scripts each. Across all these lists, listeners heard each male and each female speaker in a high- or a low-prestige job illustrating a high- and low-level performance. However, within each list, each set of listeners experienced a subset of the recordings with an equal distribution of speakers' gender (3 female, 3 male), speakers' L1 (3 Tagalog, 3 English), occupations (6 jobs), and performance levels (3 high-, 3 low-performance scenarios). The 96 listeners were assigned randomly to one of the eight versions of the interface, with 12 listeners rating each set of six audios.

Prior to completing the rating task, listeners filled out a consent form and a language background questionnaire. After reading the definition of comprehensibility, they completed

a practice block in which they evaluated this dimension using unrelated audios by three French speakers of L2 English. After providing each practice rating, listeners were directed to a new page where they received brief feedback focusing on comprehensibility (e.g., *Although you were probably able to understand what the speaker was saying, it may have taken you some effort; we expect that your rating would therefore fall around the middle of the scale*).² Listeners were then introduced to the rating scales for the remaining dimensions. In the main rating task, which included six scenarios presented in randomized order, listeners first saw a brief description contextualizing each situation (e.g., *You have just phoned an electronics store to ask about the best tablet for your child. The person who received your call left the following voicemail message for you*). This was followed by a recording which listeners clicked on to initiate playback. After hearing the audio for the first time, they proceeded to rate the speaker for comprehensibility and accentedness. They then listened to the audio a second time and provided impressions concerning the speakers' competence and the extent to which they wished to be treated as described in the script. No repeated playback was allowed after the first or the second listening. Each listener received a \$30 gift card for participating.

4.4. Data Analysis

All ratings were first checked for internal consistency, using two-way, consistency, average-measure intraclass correlations (ICCs) computed in the psych package (version 2.0.9, Revelle, 2021) in R (version 4.0.2; R Core Team, 2020). The ICC values for the full dataset were sufficiently high for the ratings of speaker competence (.84), treatment preference (.94), and comprehensibility (.84). We fitted generalized linear mixed-effects models using the lme4 package in R (Bates, Mächler, Bolker, & Walker, 2015). The three target ratings were transformed to a proportion because each metric was an integer with fixed upper and lower bounds (0–100), and a binomial distribution was used with a logit link function.³ For each model, we conducted 100,000 iterations using the BOBYQA optimizer, setting the number of

adaptive Gauss-Hermite quadrature points (nAGQ) to 0. Each model used one of the three ratings as the response variable. Listener and speaker served as random intercepts. We ran a series of models and compared them using ANOVAs to explore whether the best models might include additional random slopes. The following variables served as fixed-effects predictors: (a) speaker L1 (English or Tagalog), (b) speaker gender (female or male), (c) accentedness rating (continuous, included to capture between-speaker variation in foreign accent), (d) job prestige (high or low), (e) professional performance level (high or low), (f) interaction term between speaker L1 and accentedness rating, (g) interaction term between speaker L1 and job prestige, (h) interaction term between speaker L1 and performance level, and (i) three-way interaction of speaker L1, performance level, and job prestige. For all ratings, the best model included speaker gender as a random slope and interaction of speaker L1, performance level, and job prestige as a fixed effect.

All models also included two fixed effects as control covariates capturing listener background characteristics: number of languages listeners reported speaking (1–5) and familiarity with accented speech (1–9 scalar rating). These variables were considered particularly relevant because they have been shown to influence listener-based assessments of L2 speech. For instance, listeners' evaluations of L2 speakers differ for listeners who are monolingual, bilingual, or multilingual (Saito & Shintani, 2016), and those reporting various degrees of familiarity with L2 speech (Kennedy & Trofimovich, 2008). In light of these findings, we reasoned that listeners might evaluate various job performances more or less positively depending on their background profile. Because our goal was to compare listener evaluations as a function of speakers' L1, job prestige, and performance level, it was therefore important to control these individual differences across listeners.

Although the entire dataset was based on a total of 576 observations (96 listeners \times 6 speakers/scripts) per response variable, the lme4 function employed row-wise deletion for

missing values from either predictor or response variables. In the end, there were 551 observations for the rating of competence, 556 observations for the rating of treatment preference, and 564 observations for the rating of comprehensibility. To test the statistical significance of each parameter, we checked p values computed by the lme4 package, but also examined 95% confidence intervals (CIs) to ascertain that the interval does not cross zero. To explore statistically significant interaction effects, post hoc comparisons were performed using the emmeans package with a Tukey adjustment for p values (version 1.7.5; Lenth, 2020).

5. Results

The research question asked whether listeners' professional (competence), experiential (treatment preference), and linguistic (comprehensibility) reactions to speakers are associated with speakers' L1 (English vs. Tagalog), job prestige (high vs. low), and performance level (high vs. low), while also accounting for between-speaker variation in pronunciation through accentedness ratings. Among the three measures (summarized descriptively in Appendix B), the ratings of competence were strongly associated with those of treatment preference for English ($r = .72$) and Tagalog ($r = .71$) speakers, but these two measures showed either no relationship with English speakers' comprehensibility and accentedness ($r = .06-.15$) or only weak links with Tagalog speakers' comprehensibility and accentedness ($r = .16-.30$). In turn, comprehensibility and accentedness, as partially overlapping linguistic dimensions (Derwing & Munro, 2015), showed medium-strength relationships both for English ($r = .40$) and Tagalog ($r = .48$) speakers, based on field-specific benchmarks (Plonsky & Oswald, 2014).

5.1. Competence

As summarized in Table 1 and illustrated graphically in Figure 1, for the ratings of speaker competence, there were significant effects for speaker L1, accentedness rating, job

prestige, and performance level. Listeners rated English speakers as more competent than Tagalog speakers. Listeners also evaluated speakers who speak with a weaker foreign accent as more competent than those whose accent is stronger. Finally, listeners rated all speakers (regardless of their L1) as more competent in high-prestige jobs and in high-performance scenarios, compared to low-prestige jobs and low-performance scenarios. Neither of the two control covariates or speakers' gender emerged as significant predictors of the ratings of speaker competence.

Table 1 *Summary of Mixed-Effects Model for Speaker Competence*

Parameter	Estimate	SE	95% CI	z	p
(Intercept)	5.69	0.85	[4.12, 7.30]	6.71	< .001
Speaker L1 (Tagalog vs. English)	-3.22	0.35	[-4.01, -2.35]	-9.06	< .001
Speaker gender (male vs. female)	0.24	0.29	[-0.41, 0.81]	0.84	.401
Accentedness (heavy vs. weak)	-0.02	0.01	[-0.03, -0.01]	-7.20	< .001
Prestige (low vs. high)	-1.42	0.08	[-1.60, -1.24]	-17.27	< .001
Performance (low vs. high)	-3.89	0.08	[-4.05, -3.72]	-45.76	< .001
Speaker L1 × Accentedness	0.02	0.01	[0.01, 0.03]	6.64	< .001
Speaker L1 × Prestige	0.80	0.10	[0.62, 0.99]	8.26	< .001
Speaker L1 × Skill	2.01	0.12	[1.76, 2.25]	16.19	< .001
Speaker L1 × Prestige × Skill	-0.99	0.12	[-1.23, -0.74]	-8.03	< .001
Number of languages	0.01	0.15	[-0.27, 0.32]	0.10	.918
Familiarity with accented speech	0.01	0.09	[-0.15, 0.20]	0.13	.894

The effect for speaker L1 was qualified by an interaction with accentedness rating, where English speakers were rated as more competent than Tagalog speakers when the speaker's foreign accent was stronger, $z = 5.94$, $p < .001$, $OR = 7.13$, 95% CI [3.05, 16.68],

but not when the speaker's accent was weaker, $z = 0.61$, $p = .931$, $OR = 1.14$, 95% CI [0.65, 2.00]. The effect for job prestige was qualified by an interaction with speaker L1, whereby English speakers were rated as more competent than Tagalog speakers in high-prestige positions, $z = 3.44$, $p = .003$, $OR = 2.16$, 95% CI [1.22, 3.86], but not in low-prestige positions where the ratings for English and Tagalog speakers were similar, $z = 2.08$, $p = .157$, $OR = 1.59$, 95% CI [0.90, 2.83]. The effect of job performance was similarly qualified by an interaction with speaker L1, such that English speakers were rated as more competent than Tagalog speakers in high-performance scenarios, $z = 6.03$, $p < .001$, $OR = 3.96$, 95% CI [2.20, 7.13], but not in low-performance scenarios for which English and Tagalog speakers received comparable ratings, $z = -0.62$, $p = .927$, $OR = 0.87$, 95% CI [0.49, 1.55]. Finally, a significant three-way interaction was driven by the competence ratings in low-performance scenarios, where both English and Tagalog speakers received higher competence ratings in low- versus high-prestige jobs, but the magnitude of this difference was greater for English speakers, $z = -7.93$, $p < .001$, $OR = 0.60$, 95% CI [0.49, 0.73], than for Tagalog speakers, $z = -5.08$, $p < .001$, $OR = 0.72$, 95% CI [0.60, 0.88].

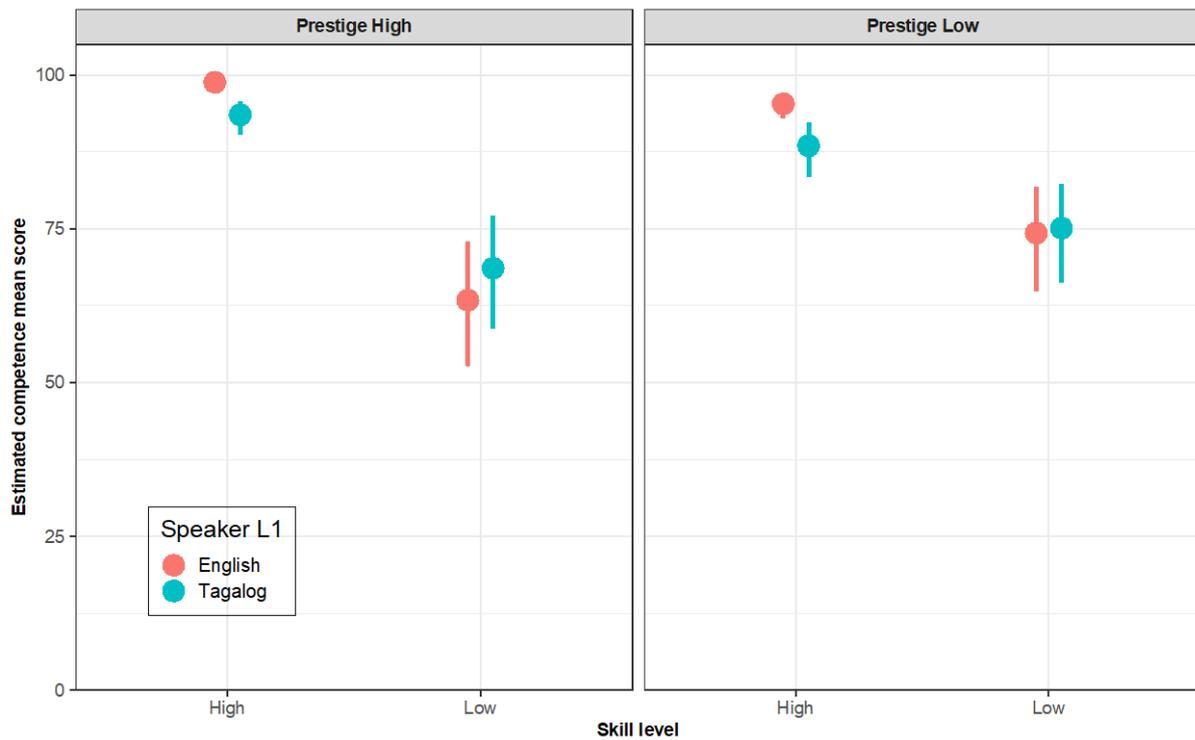


Figure 1. Estimated means for speaker competence as a function of speaker L1, performance (skill) level, and job prestige. Whiskers around estimated means enclose 95% CIs.

5.2. Treatment Preference

As shown in Table 2 and illustrated in Figure 2, for the ratings of listeners' preference to be treated as described in the script, there were significant effects for speaker L1, accentedness rating, job prestige, and performance level. Listeners gave higher preference ratings to English than Tagalog speakers and gave higher ratings to speakers who speak with a weaker foreign accent than to those whose accent is stronger. Listeners also assigned higher preference ratings to all speakers (regardless of their L1) in high-prestige jobs and in high-level performances. No control covariate, including speakers' gender, was associated with the ratings of treatment preference.

Table 2 *Summary of Mixed-Effects Model for Listener Treatment Preference*

Parameter	Estimate	SE	95% CI	<i>z</i>	<i>p</i>
(Intercept)	7.17	0.91	[5.25, 9.20]	7.88	< .001
Speaker L1 (Tagalog vs. English)	-3.56	0.38	[-4.37, -2.74]	-9.45	< .001
Speaker gender (male vs. female)	-0.20	0.30	[-0.78, 0.38]	-0.68	.497
Accentedness (heavy vs. weak)	-0.03	0.003	[-0.04, -0.03]	-12.88	< .001
Prestige (low vs. high)	-0.83	0.08	[-1.00, -0.65]	-9.83	< .001
Performance (low vs. high)	-4.37	0.09	[-4.52, -4.16]	-49.22	< .001
Speaker L1 × Accentedness	0.04	0.01	[0.03, 0.05]	13.21	< .001
Speaker L1 × Prestige	-0.23	0.11	[-0.45, 0.02]	-2.16	.031
Speaker L1 × Skill	-0.14	0.14	[-0.42, 0.15]	-0.98	.326
Speaker L1 × Prestige × Skill	1.26	0.13	[0.95, 1.52]	9.49	< .001
Number of languages	0.08	0.16	[-0.20, 0.38]	0.50	.618
Familiarity with accented speech	-0.01	0.09	[-0.19, 0.18]	-0.09	.927

The effect for speaker L1 was qualified by an interaction with accentedness rating, where English speakers received higher treatment preference ratings than Tagalog speakers when the speaker's foreign accent was stronger, $z = 9.23$, $p < .001$, $OR = 25.45$, 95% CI [10.34, 62.65], but not when the speaker's accent was weaker, $z = -1.96$, $p = .200$, $OR = 0.61$, 95% CI [0.32, 1.17]. The effect of job prestige was qualified by an interaction with speaker L1, whereby English speakers received higher treatment preference ratings than Tagalog speakers in high-prestige jobs, $z = 2.81$, $p = .025$, $OR = 2.05$, 95% CI [1.06, 3.97], but not in low-prestige jobs where English and Tagalog speakers' ratings were similar, $z = 1.25$, $p = .591$, $OR = 1.37$, 95% CI [0.72, 2.65]. Finally, a significant three-way interaction was driven by the treatment preference ratings given to Tagalog speakers (rather than English speakers)

in low-performance scenarios, where Tagalog speakers received higher ratings in low- versus high-prestige jobs, $z = -15.72$, $p < .001$, $OR = 0.367$, 95% CI [0.30, 0.44].

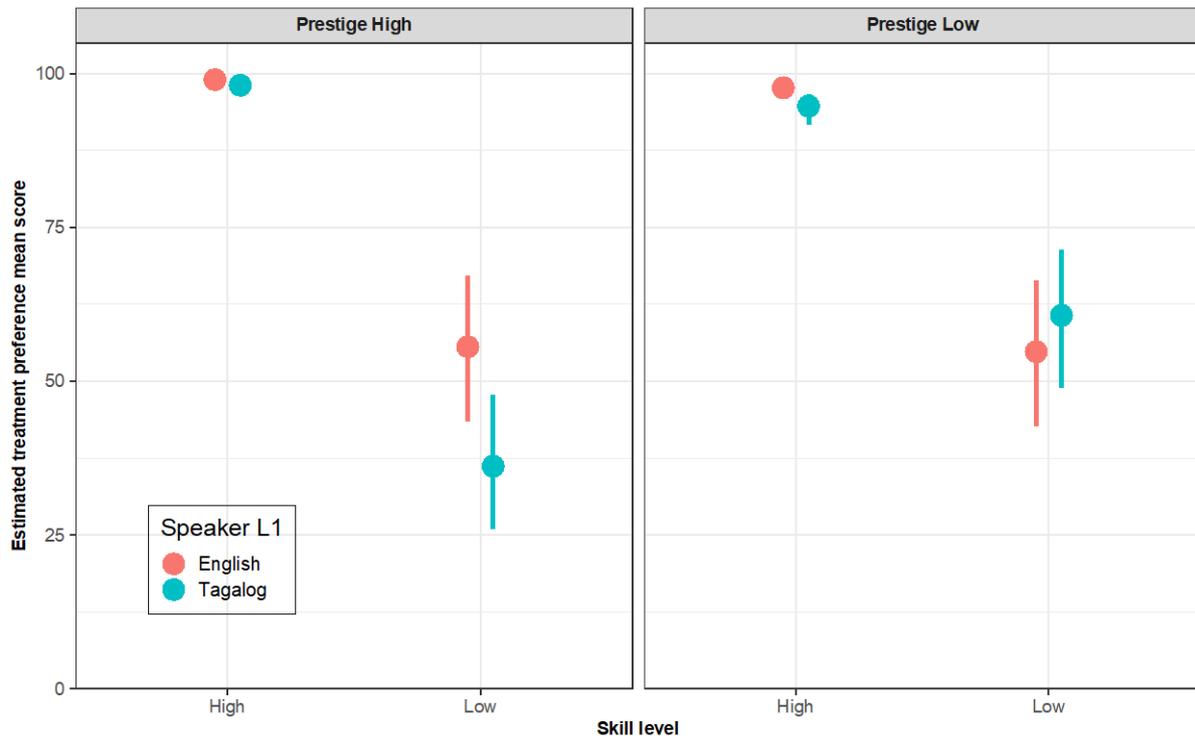


Figure 2. Estimated means for treatment preference as a function of speaker L1, performance (skill) level, and job prestige. Whiskers around estimated means enclose 95% CIs

5.3. Comprehensibility

As shown in Table 3 and illustrated in Figure 3, for the ratings of comprehensibility, there were significant effects for speaker L1, accentedness rating, job prestige, and performance level. Listeners rated Tagalog speakers as less comprehensible than English speakers. They rated speakers with weaker accents as more comprehensible than those whose accent is stronger. Listeners also rated all speakers (regardless of their L1) as more comprehensible in high-prestige positions and in high-performance scenarios. Again, neither control covariate or speakers' gender was associated with comprehensibility ratings.

Table 3 *Summary of Mixed-Effects Model for Speaker Comprehensibility*

Parameter	Estimate	SE	95% CI	<i>z</i>	<i>p</i>
(Intercept)	0.33	0.90	[-1.43, 2.20]	0.37	.709
Speaker L1 (Tagalog vs. English)	1.53	0.56	[0.34, 2.65]	2.74	.006
Speaker gender (male vs. female)	0.39	0.45	[-0.49, 1.20]	0.86	.392
Accentedness (heavy vs. weak)	0.05	0.01	[0.04, 0.06]	14.83	< .001
Prestige (low vs. high)	-0.84	0.15	[-1.10, -0.58]	-5.53	< .001
Performance (low vs. high)	-1.39	0.14	[-1.70, -1.09]	-9.74	< .001
Speaker L1 × Accentedness	-0.03	0.01	[-0.04, -0.02]	-8.06	< .001
Speaker L1 × Prestige	0.04	0.16	[-0.22, 0.34]	0.23	.817
Speaker L1 × Skill	0.34	0.18	[-0.02, 0.70]	1.85	.065
Speaker L1 × Prestige × Skill	-0.84	0.20	[-1.25, -0.45]	-4.13	< .001
Number of languages	-0.03	0.14	[-0.30, 0.25]	-0.19	.850
Familiarity with accented speech	-0.06	0.08	[-0.23, 0.09]	-0.72	.472

The effect of speaker L1 was qualified by an interaction with accentedness rating, where the rated difference in comprehensibility between English and Tagalog speakers (in favor of English speakers) was more pronounced when the speaker’s accent was stronger, $z = 3.98$, $p < .001$, $OR = 5.38$, 95% CI [1.82, 15.93], than when the speaker’s foreign accent was weaker, $z = -2.57$, $p = .05$, $OR = 0.26$, 95% CI [0.07, 1.00]. Finally, a significant three-way interaction between speaker L1, prestige, and performance level was driven by the ratings in low-performance scenarios, where both English and Tagalog speakers were evaluated as sounding more comprehensible in low- versus high-prestige jobs, but the magnitude of this difference was greater for English speakers, $z = -9.82$, $p < .001$, $OR = 0.31$, 95% CI [0.22, 0.45], than for Tagalog speakers, $z = -4.97$, $p < .001$, $OR = 0.71$, 95% CI [0.58, 0.87].

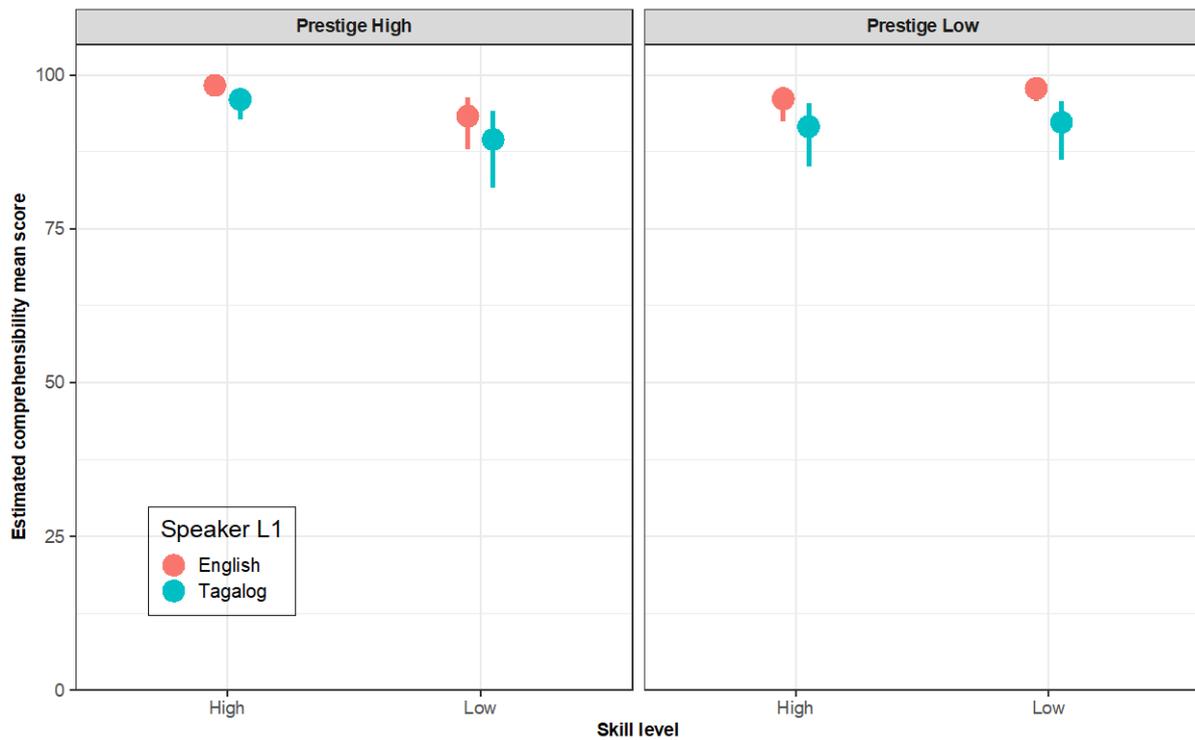


Figure 3. Estimated means for speaker comprehensibility as a function of speaker L1, performance (skill) level, and job prestige. Whiskers around estimated means enclose 95% CIs

6. Discussion

6.1. Summary of Findings

In this study, which was inspired by research on accentism in workplace-relevant contexts, we investigated how listeners evaluate the performance of native and L2 speakers of English in simulated job scenarios that illustrate high versus low professional performances in high- versus low-prestige occupations. Our aim was to examine accent bias through several listener-rated metrics, including competence, treatment preference, and comprehensibility. Broadly speaking, our findings aligned with prior work on listener reactions to L2 speakers (Baquiran & Nicoladis, 2020; Timming, 2017), inasmuch as native English speakers were generally perceived as more competent and comprehensible and indeed received higher preference ratings than Tagalog speakers in almost all job prestige and

performance scenarios. Listeners' evaluations also depended on job status (Goyder et al., 2003; Neiterman & Bourgeault, 2015), in that speakers in high-prestige jobs (doctor, lawyer, professor) tended to receive higher ratings than those in low-prestige jobs (cleaner, salesperson, server). Lastly, listener assessments reflected professional attributes of the speaking performances, in that high-performance scenarios were judged more favorably than low-performance scenarios. Comparatively speaking, speakers' L1 and professional performance level had a far greater effect on the ratings than job prestige (see Figures 1–3).

These general patterns aside, when considered at a finer-grained level, our findings provided a nuanced picture of how listeners assess job-relevant speaking performances. For L2 speakers in particular, listener evaluations appeared to depend on job prestige and professional performance level. When job prestige was low, Tagalog speakers were perceived just as competent as and received similar preference ratings to English speakers. When the speaker's professional performance was low, Tagalog speakers received similar competence ratings to English speakers. These general patterns become even more nuanced when we consider the effects of both job prestige and performance level (discussed below) and factor in the strength of the speaker's accent. With respect to L2 accent, in particular, the magnitude of rating differences for speaker competence, treatment preference, and comprehensibility in favor of English speakers was greater when the speaker's foreign accent was stronger than when the speaker's accent was weaker, which implies a gradient (rather than categorical) influence of foreign accent on speaker evaluations. These findings, which provide novel insights to existing literature, were obtained after we statistically controlled two variables capturing individual differences in listeners' monolingual versus multilingual status and their familiarity with accented speech (see Tables 2–4).

6.2. Competence and Treatment Preference

The ratings of competence and treatment preference generally patterned together in describing listeners' reactions to speakers. The association between these measures was strong ($r = .71-.72$), with around 50% of variance in common, which is consistent with research in business and marketing, where perceived competence of workers in various retail and service positions, along with their behavior toward customers, predicts various measures of customer satisfaction, preference, loyalty, service quality, and trust (Darian et al., 2001; Hennig-Thurau, 2004; Walsh et al., 2012). In essence, listeners may have relied on their assessments of speaker competence in expressing their treatment preference.

As shown through complex three-way interactions involving speaker L1, job prestige, and performance level (see Figures 1 and 2), despite their general bias in favor of native English speakers, listeners upgraded Tagalog speakers in their ratings of competence and treatment preference when these speakers performed a low-prestige job and simultaneously illustrated a low-level performance. On the one hand, Filipinos are valued for their expertise and work ethic (Coloma et al., 2012; Lightman et al., 2021), which may have factored into listeners' evaluations, although this interpretation requires that listeners reliably identify the ethnolinguistic origin of the speaker, which is a challenging task for many listeners (Dragojevic & Goatley-Soan, 2022). On the other hand, Filipinos often occupy lower-status positions (Semyonov & Gorodzeisky, 2004), similar to other non-preferred groups (De La Zerda & Hopper, 1979; Iheduru-Anderson, 2020). Listeners may have thus perpetuated this idea by favoring Tagalog speakers in the scenarios where job prestige and performance expectations were low.

However, an undifferentiated bias against Tagalog speakers was clearly not present in this dataset, because listeners similarly upgraded English speakers in their competence ratings when these speakers performed low-prestige jobs and illustrated low-level

performances. In essence, listeners rewarded both English and Tagalog speakers in their evaluations when the speakers were cast in a particularly unfavorable light—performing a low-prestige job while also demonstrating a low-level professional performance. It may be that listeners empathized with the speakers who found themselves in this particularly unfortunate professional circumstance. Alternatively, listeners may have “compensated” for their general biases by rewarding the speakers whom they found especially lacking in ability or performance (Pantos & Perkins, 2013).

Regardless of the explanation, listener responses were highly nuanced. When job prestige and performance level were both low, it was English speakers, rather than Tagalog speakers, who received a stronger boost in competence ratings (see Figure 1). Listeners thus appeared to conflate language skill with professional competence for L2 speakers, but did so to a lesser degree when assessing native speakers. Put another way, unlike L2 speakers whose competence was likely attributed to their accented speech, native speakers got “excused” for poor performances in low-prestige jobs. In contrast, in exactly the same situation when job prestige and performance level were both low, it was Tagalog speakers, not English speakers, who received a boost in treatment preference (see Figure 2). In this case, listeners seemed to show a general acceptance to be treated by L2 speakers in low-prestige, low-performance scenarios, suggesting that they might expect L2 speakers to occupy low-prestige positions *and* to demonstrate predictably poor professional performance (Iheduru-Anderson, 2020; Kalin & Rayko, 1978).

6.3. Comprehensibility

With respect to comprehensibility, although English speakers were considered easier to understand overall, Tagalog speakers received relatively high ratings (69–88 on a 100-point scale, see Appendix B), supporting the idea that Tagalog-accented English often poses little difficulty for listeners (Dayag, 2007; Li & Chen, 2019). Nevertheless, Tagalog

speakers' comprehensibility depended on job prestige and performance level, where basically the same speakers were judged as easier to understand in high-prestige and high-performance scenarios. Given that a speaker's language, in terms of its pronunciation, lexis, and grammar, can account for up to 85% of variance in listener-rated comprehensibility (Crowther, Trofimovich, Saito, & Isaacs, 2018), it is noteworthy that comprehensibility was influenced by nonlinguistic, contextual factors (Taylor Reid, Trofimovich, & O'Brien, 2019). As with the ratings of competence and treatment preference, these factors presumably tapped into listener expectations, such that foreign-accented speakers belong in low-prestige occupations (Kalin et al., 1980; Kalin & Rayko, 1978; Ryan & Sebastian, 1980; Stewart et al., 1985) or that they rarely perform their jobs at a greater-than-expected professional level (Iheduru-Anderson, 2020). When these expectations were violated, the speakers were rewarded by being judged as more comprehensible. An even more striking contextual effect on comprehensibility occurred in low-prestige, low-performance scenarios, where both English and Tagalog speakers' comprehensibility was upgraded, similar to competence and treatment preference ratings, most likely as a consequence of listeners compensating speakers for low-level performance in dispreferred (low-status) occupations.

Finally, the comprehensibility ratings of Tagalog speakers were associated with their ratings of competence ($r = .28$) and treatment preference ($r = .30$). Although these relationships were weak (Plonsky & Oswald, 2014), they supported a previous claim that difficulty understanding a foreign-accented speaker can trigger low ratings of competence, and reciprocally, low ratings of competence can aggravate potential comprehensibility issues (Dragojevic & Goatley-Soan, 2022). Considering that no such associations obtained for native English speakers ($r = .12-.15$), even weak links between linguistic and professional assessments of L2 speakers are a source of concern. This is because L2 speakers will often find themselves in a no-win situation in workplace settings, as their accent can trigger social

biases, resulting in negative professional evaluations, which can in turn exacerbate potential comprehensibility issues that could further jeopardize their assessments.

6.4. Listener Variables as Predictors of Speaker Ratings

To capture potential effects of listeners' language experience, we included two variables as control covariates in statistical modeling: the number of languages each listener speaks (1 through 5) and listeners' familiarity with accented speech (1–9 scalar rating). Although interaction in and with speakers of other languages may be associated with ratings of L2 speech (e.g., Saito & Shintani, 2016), neither emerged as a significant predictor of any target assessment. These results are compatible with recent evidence that monolingual and multilingual listeners are comparable in the speech ratings they provide when assessing global dimensions of L2 speech (Saito, 2021), although these findings do not necessarily capture potential similarities or differences in how L1 and L2 listeners (particularly those who share a language background with the speakers being evaluated) rate L2 speaking performances. While previous studies have shown that familiarity with accented speech may have a positive impact on listener ratings (Kennedy & Trofimovich, 2008), the lack of association in the current study may be due to contextual factors. Listeners in the current study, all residents of Calgary, are likely to interact frequently with native speakers of Tagalog, who make up 6% of the city's population. Considering listeners' high exposure to Tagalog-accented English in Calgary and their generally high overall self-rated familiarity with L2-accented English (7.65 on a 9-point scale), it is perhaps unsurprising that this listener variable contributed little to explaining the ratings.

7. Limitations and Future Work

The present study is not without limitations. First, because speakers included only members of one linguistic group, it is impossible to generalize the findings to other L2 speakers, including those that belong to other non-preferred (e.g., Latin, Middle Eastern)

versus preferred (e.g., Western European) ethnolinguistic backgrounds. Similarly, a finer-grained perspective on listeners' rating behaviors would require a systematic comparison of various listener groups, such as native speakers, bilinguals and multilinguals, managerial staff, and HR specialists, and consideration of other listener variables not targeted here, including listeners' gender, professional occupation, and prior experience with accentism or discrimination.

Second, speakers' job performances were scripted and were assessed in a context that did not provide listeners with access to speakers' visual cues, such as gestures, body language, and facial expressions, which could have influenced their professional, experiential, and linguistic assessments. Third, it is possible that listener bias may be more or less pronounced depending on the rated measure, so workplace-relevant performances need to be evaluated for other dimensions, including pragmalinguistic and communicative (e.g., politeness, sensitivity to client needs). Until clarified in future work, our present interpretations concern only individual effects of job prestige and performance level for the two speaker groups targeted here. Finally, although our study yielded workplace-relevant insights, it was not conducted in a workplace setting, which limits the generalizability of our findings and calls for additional work carried out in actual employment contexts.

While we are unable to determine whether the listeners in this study expected a speaker to be less skilled or to underperform, their ratings implied that they make speech-based assumptions about the speaker's performance. Such assumptions may have important implications for the workplace. For instance, in pre-employment situations, native speakers may be particularly favored in their search for high-prestige jobs and possibly disadvantaged in low-prestige occupations. Similarly, in in-employment settings, expectations about professional performance might differ for native and L2 speakers, such that native speakers may receive less latitude when it comes to performing a job less competently whereas skilled

job performances by L2 speakers may be received with excessive surprise or incredulity. Broadly construed, future research on this topic should investigate the extent to which listeners' judgments of job performance can be affected by a task that encourages empathy with L2 speakers' experiences on the job market, on the assumption that listeners' awareness of these experiences might mitigate their bias. Given the gatekeeping function of HR professionals, we also propose that subsequent investigations take place within HR settings, including those that train future HR specialists. Such work will provide insights into current practices and will be a testing ground for potential bias-reduction programs.

8. Conclusion

In this study, we demonstrated that accent bias may be more nuanced than generally assumed in that listeners do not show an undifferentiated bias against L2 speakers. While native English speakers were assigned more favorable ratings overall, Tagalog speakers were rated similarly to English speakers in low-prestige jobs and in low-performance scenarios, and both English and Tagalog speakers were rewarded in ratings when job prestige was low and performance level was simultaneously poor. Such nuanced biases, however, are at least as damaging as broad, undifferentiated ones. Listeners likely assumed that L2 speakers might be especially suitable for low-paying, low-status jobs, particularly those that do not attract native-born applicants. Listeners furthermore demonstrated an assumption that poor professional performances might be expected of speakers employed in low-paying, low-status jobs. These assumptions are likely driven by listeners' experience seeing foreign-accented speakers in these positions or by their expectations of which jobs or which performances are associated with those speakers. Needless to say, more work is needed to understand and ultimately reduce accentism in workplace contexts.

Notes

1. We acknowledge that the high- versus low-level performance labels used here to describe scripted job performances encompass more than each speaker's professional job skill and capture a variety of performance differences, such as a speaker's job experience, quality of training, sensitivity to the needs of others, politeness, and effectiveness. We leave it up to follow-up research to disentangle the role of various specific performance indicators (e.g., a speaker's politeness, use of concrete examples in addressing a customer) in rater evaluations. In the interim, our labels of "high" versus "low" performance should be understood broadly, as representing generally better versus worse professional performance.
2. An anonymous reviewer indicated that the specificity of the feedback provided to listeners may have had an impact on the comprehensibility ratings they provided. Because this was a fully online study in which the researcher was unable to interact with listeners, we made the decision to include specific feedback to encourage them to contemplate the difference between accentedness and comprehensibility ratings.
3. Min–Max normalization was used to transform the 0–100 scale to 0–1. It preserves the shape of the original distribution and is least disruptive to information in the original data. After this transformation, all ratings of competence, treatment preference, and comprehensibility were between 0 and 1. We then fitted generalized linear mixed-effects models for each of the target ratings. Binomial distribution was chosen as the family with logit link function in this study, which is an appropriate choice when a response variable is a proportion (Baum, 2008).

References

- Almeida, S., Fernando, M., Hannif, Z., & Dharmage, S. C. (2015). Fitting the mould: The role of employer perceptions in immigrant recruitment decision-making. *International Journal of Human Resource Management*, 26(22), 2811–2832.

- <https://doi.org/10.1080/09585192.2014.1003087>
- Baquiran, C. L. C., & Nicoladis, E. (2020). A doctor's foreign accent affects perceptions of competence. *Health Communication, 35*(6), 726–730.
- <https://doi.org/10.1080/10410236.2019.1584779>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48.
- <https://doi.org/10.18637/jss.v067.i01>
- Baum, C. F. (2008). Stata tip 63: Modeling proportions. *Stata Journal, 8*(2), 299–303.
- <https://doi.org/10.1177/1536867X0800800212>
- Boersma, P., & Weenink, D. (2021). *Praat: Doing phonetics by computer* [Computer program]. Version 6.1.41, retrieved from <http://www.praat.org/>
- Brennan, E. M., & Brennan, J. S. (1981). Accent scaling and language attitudes: Reactions to Mexican American English speech. *Language and Speech, 24*(3), 207–221.
- <https://doi.org/10.1177/002383098102400301>
- Cargile, A. C., & Giles, H. (1998). Language attitudes toward varieties of English: An American-Japanese context. *Journal of Applied Communication Research, 26*(3), 338–356. <https://doi.org/10.1080/00909889809365511>
- Coloma, R. S., McElhinny, B., Tungohan, E., Catungal, J. P., & Davidson, L. M. (Eds.). (2012). *Filipinos in Canada: Disturbing invisibility*. Toronto, CA: Toronto Univeristy Press.
- Crowther, D., Trofimovich, P., Isaacs, T., & Saito, K. (2018). Linguistic dimensions of L2 accentedness and comprehensibility vary across speaking tasks. *Studies in Second Language Acquisition, 40*(2), 443–457. <https://doi.org/10.1017/S027226311700016X>
- Darian, J. C., Tucci, L. A., & Wiman, A. R. (2001). Perceived salesperson service attributes and retail patronage intentions. *International Journal of Retail Distribution*

- Management*, 29(5), 205–213. <https://doi.org/10.1108/09590550110390986>
- Dávila, A., Bohara, A. K., & Saenz, R. (1993). Accent penalties and the earnings of Mexican Americans. *Social Science Quarterly*, 74(4), 902–916.
- Dayag, D. T. (2007). Exploring the intelligibility of Philippine English. *Asian Englishes*, 10(1), 4–23. <https://doi.org/10.1080/13488678.2007.10801197>
- de Castro, A. B., Gee, G. C., & Takeuchi, D. T. (2008). Workplace discrimination and health among Filipinos in the United States. *American Journal of Public Health*, 98(3), 520–526. <https://doi.org/10.2105/AJPH.2007.110163>
- De La Zerda, N., & Hopper, R. (1979). Employment interviewers' reactions to Mexican American speech. *Communication Monographs*, 46(2), 126–134. <https://doi.org/10.1080/03637757909375998>
- Deprez-Sims, A.-S., & Morris, S. B. (2010). Accents in the workplace: Their effects during a job interview. *International Journal of Psychology*, 45(6), 417–426. <https://doi.org/10.1080/00207594.2010.499950>
- Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. Amsterdam, NL: John Benjamins.
- Dragojevic, M., Gasiorek, J., & Giles, H. (2016). Communication accommodation theory. In C. R. Berger & M. E. Roloff (Eds.), *The international encyclopedia of interpersonal communication* (pp. 1–20). Hoboken, NJ: Wiley. <https://doi.org/10.1002/9781118540190.wbeic006>
- Dragojevic, M., & Goatley-Soan, S. (2022). Americans' attitudes toward foreign accents: Evaluative hierarchies and underlying processes. *Journal of Multilingual and Multicultural Development*, 43(2), 167–181. <https://doi.org/10.1080/01434632.2020.1735402>
- Giles, H., & Watson, B. M. (Eds.). (2013). *The social meanings of language, dialect, and*

- accent: International perspectives on speech styles*. Bern, Switzerland: Peter Lang.
- Go, S. P. (1998). Towards the 21st century: Whither Philippine labor migration. In B. V. Cariño (Ed.), *Filipino workers on the move: Trends, dilemmas and policy options* (pp. 9–44). Philippine Migration Research Network.
- Goyder, J., Guppy, N., & Thompson, M. (2003). The allocation of male and female occupational prestige in an Ontario urban area: A quarter-century replication. *Canadian Review of Sociology/Revue Canadienne de Sociologie*, 40(4), 417–439.
<https://doi.org/10.1111/j.1755-618X.2003.tb00255.x>
- Hamilton, D., Goldsmith, A. H., & Darity, W. (2008). Measuring the wage costs of limited English: Issues with using interviewer versus self-reports in determining Latino wages. *Hispanic Journal of Behavioral Sciences*, 30(3), 257–279.
<https://doi.org/10.1177/0739986308320470>
- Hauser, R. M., & Warren, J. R. (1997). Socioeconomic indexes for occupations: A review, update, and critique. *Sociological Methodology*, 27(1), 177–298.
<https://doi.org/10.1111/1467-9531.271028>
- Hennig-Thurau, T. (2004). Customer orientation of service employees: Its impact on customer satisfaction, commitment, and retention. *International Journal of Service Industry Management*, 15(5), 460–478. <https://doi.org/10.1108/09564230410564939>
- Hosoda, M., Stone-Romero, E. F., & Walter, J. N. (2007). Listeners' cognitive and affective reactions to English speakers with standard American English and Asian accents. *Perceptual and Motor Skills*, 104(1), 307–326. <https://doi.org/10.2466/PMS.104.1.307-326>
- Iheduru-Anderson, K. (2020). Accent bias: A barrier to Black African-born nurses seeking managerial and faculty positions in the United States. *Nursing Inquiry*, 27(4), 1–15.
<https://doi.org/10.1111/nin.12355>

- Kalin, R., & Rayko, D. S. (1978). Discrimination in evaluative judgments against foreign-accented job candidates. *Psychological Reports*, 43(3), 1203–1209.
<https://doi.org/10.2466/pr0.1978.43.3f.1203>
- Kalin, R., Rayko, D. S., & Love, N. (1980). The perception and evaluation of job candidates with four different ethnic accents. In H. Giles, W. P. Robinson, & P. M. Smith (Eds.), *Language: Social psychological perspectives* (pp. 197–202). Oxford, UK: Pergamon Press. <https://doi.org/10.1016/B978-0-08-024696-3.50034-X>
- Kennedy, S., & Trofimovich, P. (2008). Intelligibility, comprehensibility, and accentedness of L2 speech: The role of listener experience and semantic context. *Canadian Modern Language Review*, 64(3), 459–490. <https://doi.org/10.3138/cmlr.64.3.459>
- Lenth, R. (2020). emmeans: Estimated marginal means. R package version 1.7.5.
<https://CRAN.R-project.org/package=emmeans>
- Li, H. P., & Chen, H. C. (2019). Intelligibility and comprehensibility of the Filipino English accent to Hong Kong English speakers. *The Southeast Asian Journal of English Language Studies*, 25(1), 23–42. <https://doi.org/10.17576/3L-2019-2501-02>
- Lightman, N., Banerjee, R., Tungohan, E., de Leon, C., & Kelly, P. (2021). An intersectional pathway penalty: Filipina immigrant women inside and outside Canada’s Live-In Caregiver Program. *International Migration*, 1–20. <https://doi.org/10.1111/imig.12851>
- Lindemann, S. (2003). Koreans, Chinese or Indians? Attitudes and ideologies about non-native English speakers in the United States. *Journal of Sociolinguistics*, 7(3), 348–364.
<https://doi.org/10.1111/1467-9481.00228>
- Lippi-Green, R. (2012). *English with an accent: Language, ideology and discrimination in the United States* (2nd ed.). New York, NY: Routledge.
- Mai, R., & Hoffmann, S. (2014). Accents in business communication: An integrative model and propositions for future research. *Journal of Consumer Psychology*, 24(1), 137–158.

<https://doi.org/10.1016/j.jcps.2013.09.004>

Moyer, A. (2013). *Foreign accent: The phenomenon of non-native speech*. Cambridge, UK: Cambridge University Press.

Neiterman, E., & Bourgeault, I. L. (2015). The shield of professional status: Comparing internationally educated nurses' and international medical graduates' experiences of discrimination. *Health (United Kingdom)*, *19*(6), 615–634.

<https://doi.org/10.1177/1363459314567788>

Nelson Jr., L. R., Signorella, M. L., & Botti, K. G. (2016). Accent, gender, and perceived competence. *Hispanic Journal of Behavioral Sciences*, *38*(2), 166–185.

<https://doi.org/10.1177/0739986316632319>

Pantos, A. J., & Perkins, A. W. (2013). Measuring implicit and explicit attitudes toward foreign accented speech. *Journal of Language and Social Psychology*, *32*(1), 3–20.

<https://doi.org/10.1177/0261927X12463005>

Plonsky, L., & Oswald, F. L. (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language Learning*, *64*(4), 878–912. <https://doi.org/10.1111/lang.12079>

Ramjattan, V. A. (2019). Racializing the problem of and solution to foreign accent in business. *Applied Linguistics Review*, 1–18. <https://doi.org/10.1515/applirev-2019-0058>

Reitz, J. G., & Sklar, S. M. (1997). Culture, race, and the economic assimilation of immigrants. *Sociological Forum*, *12*(2), 233–277.

<https://doi.org/10.1023/A:1024649916361>

R Core Team (2020). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org>

Revelle, W. (2021). *psych: Procedures for psychological, psychometric, and personality research*. Northwestern University, Evanston, Illinois. R package version 2.0.9,

<https://CRAN.R-project.org/package=psych>

- Roberts, C. (2021). *Linguistic penalties and the job interview*. Bristol, UK: Equinox.
- Ryan, E. B., & Sebastian, R. J. (1980). The effects of speech style and social class background on social judgements of speakers. *British Journal of Social and Clinical Psychology, 19*(3), 229–233. <https://doi.org/10.1111/j.2044-8260.1980.tb00348.x>
- Saito, K. (2021). What characterizes comprehensible and native-like pronunciation among English-as-a-second-language speakers? Meta-analyses of phonological, rater, and instructional factors. *TESOL Quarterly, 55*(3), 866–900. <https://doi.org/10.1002/tesq.3027>
- Saito, K., & Shintani, N. (2016). Do native speakers of North American and Singapore English differentially perceive comprehensibility in second language speech? *TESOL Quarterly, 50*(2), 421–446. <https://doi.org/10.1002/tesq.234>
- Segrest Purkiss, S. L., Perrewé, P. L., Gillespie, T. L., Mayes, B. T., & Ferris, G. R. (2006). Implicit sources of bias in employment interview judgments and decisions. *Organizational Behavior and Human Decision Processes, 101*(2), 152–167. <https://doi.org/10.1016/j.obhdp.2006.06.005>
- Semyonov, M., & Gorodzeisky, A. (2004). Occupational destinations and economic mobility of Filipino overseas workers. *International Migration Review, 38*(1), 5–25. <https://doi.org/10.1111/j.1747-7379.2004.tb00186.x>
- Statistics Canada. (2017). Calgary, CY [Census subdivision], Alberta and Canada [Country] (table). Census Profile. 2016 Census. Statistics Canada Catalogue no. 98-316-X2016001. Ottawa. Released November 29, 2017. <https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/index.cfm?Lang=E> (accessed March 7, 2022).
- Stewart, M. A., Ryan, E. B., & Giles, H. (1985). Accent and social class effects on status and solidarity evaluations. *Personality and Social Psychology Bulletin, 11*(1), 98–105. <https://doi.org/10.1177/0146167285111009>

- Taylor Reid, K., Trofimovich, P., & O'Brien, M. G. (2019). Social attitudes and speech ratings: Effects of positive and negative bias on multiage listeners' judgments of second language speech. *Studies in Second Language Acquisition*, *41*(2), 419–442. <https://doi.org/10.1017/S0272263118000244>
- Timming, A. R. (2017). The effect of foreign accent on employability: A study of the aural dimensions of aesthetic labour in customer-facing and non-customer-facing jobs. *Work, Employment and Society*, *31*(3), 409–428. <https://doi.org/10.1177/0950017016630260>
- Walsh, G., Gouthier, M., Gremler, D. D., & Brach, S. (2012). What the eye does not see, the mind cannot reject: Can call center location explain differences in customer evaluations? *International Business Review*, *21*(5), 957–967. <https://doi.org/10.1016/j.ibusrev.2011.11.002>
- Zschirnt, E., & Ruedin, D. (2016). Ethnic discrimination in hiring decisions: A meta-analysis of correspondence tests 1990–2015. *Journal of Ethnic and Migration Studies*, *42*(7), 1115–1134. <https://doi.org/10.1080/1369183X.2015.1133279>

Appendix A

Scripts of Evaluated Recordings

Job	Low-level performance	High-level performance
Salesperson	<p>You have just phoned an electronics store to ask about the best tablet for your child. The person who received your call left the following voicemail message for you.</p> <p>Hi! This is Alexis from the electronics store. Your message was transferred to me, and I understand that you are looking for a new tablet for your child to do their homework. There are a lot of options out there, and I can't make your decision for you. You really need to look around, do your own research, and make the decision that works for you. Think about what you want, and buy a tablet that has those features. Some technology websites might help you make your decision. Once you've checked things out and decided, you can phone or come in to make your purchase. Please make sure to purchase it from me since I helped you.</p>	<p>Hi! This is Alexis from the electronics store. Your message was transferred to me. I understand that you are looking for a new tablet for your child to do their homework. Before I make a recommendation, I'd like to ask you to provide me with some information. First of all, I'll need to know how old your child is and what kinds of things they'll need the tablet for. Another important consideration is your own technology preference (that is, Apple or PC) and whether it matters if your child's device aligns with your own. Finally, if you could provide me with a price range, I can give you some good options by the end of today.</p>
Cleaner	<p>You would like to hire a cleaning service for your workplace. Remy, who works at the cleaning service, has left the following voicemail message for you.</p> <p>Hi! My name is Remy, and I work in client relations for the cleaning service. I got your message and will give you an estimate, but you need to give me some details first. What do you need to have cleaned? Where is it? How often do you want us in? And how many cleaners do you want to have each time? You need to give me this information before I can tell you anything. What I <u>can</u> say is that the more time we spend there, the more it'll cost you. I'm pretty busy right now, so it will probably take me a few weeks to get back to you with an estimate.</p>	<p>Hi! My name is Remy, and I work in client relations for the cleaning service. I understand that you'd like an estimate. Before I provide you with one, I'd like to ask you to please provide us with some additional information. First of all, it's important to know how often you'd like us to clean. If we clean more frequently, it will take us less time and cost less each time we clean. Secondly, please let me know the total number of offices, washrooms and kitchens you need us to clean, since we charge different amounts for each. I'll get back to you in about a day to provide you with an estimate.</p>
Server	<p>You have just sat down at a table at a restaurant with eight of your friends. Your server says the following as you sit down.</p>	

Good evening! Welcome to Mendoza's Restaurant. My name is Ira, and I'll be your server. Since we've been around and serving family-style meals for so long, I'm sure that you you've heard how the restaurant works. Everything is listed on the menu. Order as much food as you think you'll need. Most people say that they like our food. I can't really recommend anything, because I haven't tried most of the dishes. You might want to look around the restaurant to see what others have ordered. I'll leave the menus here on the table, and I'll take your order when I get back.

Good evening! Welcome to Mendoza's Restaurant. My name is Ira, and I'll be your server. I understand this is your first time dining with us, so I'll tell you about how things work. We are a family-style restaurant, and all of our meals feed up to three people. Because there are nine of you today, I'd recommend that you order three meals along with a soup and salad to share. The soup and salad are complimentary. I'd be happy to recommend some of my favourites, but everything we have is really delicious. I'll bring out some complimentary bread in just a moment, and I'll answer any questions then.

Doctor	<p>You have just had some lab tests done. Good morning! This is Dr. Santos. I just got your lab results. I'm really busy, so I'll have to be quick. Your CBC shows that your RBCs are 4.7 million cells per microliter, and your WBCs are 5,000 cells per microliter. Your LFTs are normal, and your total Bilirubin levels are within normal ranges. The blood culture test looked pretty bad, and it showed that you have an infection. Please call back and ask the nurse for your prescription. We'll probably need to try a few different antibiotics to clear it up. You can check out WebMD if you don't understand anything I've said. We can schedule your procedure sometime in the next year or so.</p>	<p>Dr. Santos, your family doctor, has left the following voicemail to let you know the results. Good morning! This is Dr. Santos. I've just received the results of your lab tests, and the news is good. Your complete blood count looks normal. I don't see any signs of clotting issues. Your liver enzymes are also right where they need to be. It looks like your liver is functioning normally; however, the results of the blood culture test show that you might have a bladder infection. I've called in a prescription to your local pharmacy for some antibiotics, and those should clear up your infection. Please feel free to phone back and ask for me if you have questions. We should be able to schedule your procedure for the end of next week.</p>
Lawyer	<p>You have hired a lawyer named Angel to assist you with a case. They have left you the following voicemail message. Hello! This is Angel, the lawyer assigned to your case. I came across your files under some things on my desk and will start working on the paperwork before too long. There are bits of information I can't find yet, so I don't know how long it will take. Nonetheless, it would be good</p>	<p>Hello! This is Angel, the lawyer assigned to your case. My paralegal has shared all of your information with me, and I have read through it all very carefully. On the basis of this information, I have prepared all of the paperwork that you will need to sign. In addition to this, I have put</p>

if you could come in to the office so that I can start on the documents you'll need to sign. At the meeting we can talk about what you think I should do next. You can let me know what kind of other evidence you have. Please phone my assistant to set up an appointment.

together a checklist for moving forward with your case. It includes the precise list of documents and other evidence we will need as well as a clear timeline for moving ahead. Please phone or email my assistant to set up an appointment at which you can sign the paperwork and we can map out your next steps.

Professor

You submitted an assignment in one of your classes online. Your professor has provided a recording with the following feedback.

Hi, Quinn! Thanks for submitting your first essay. I just looked at it really briefly, and I'm pretty disappointed with what you submitted. It needs a lot of work. I didn't understand most of what you wrote. It seems like you did not spend enough time on it, and you clearly have a lot to learn. Obviously, I really didn't like it. There are some good websites that provide advice on writing. Take a look at those before you submit the next assignment. Maybe it will meet my expectations more than this current assignment did.

Hi, Quinn! Thanks for submitting your first essay. I had the opportunity to read through it, and I think that you have clearly addressed most of the main points outlined in the assignment. In addition, you have organized and edited the paper well and have provided a strong rationale for the first two arguments. Your third argument was a bit less developed. I recommend that you cite previous research and provide data to support the argument. Your conclusion is less effective because it is personal. Given that this is a persuasive essay, please conclude it in a more objective manner.

Appendix B

Descriptive Statistics for Listener Ratings

Means (Standard Deviations) for Ratings of L1 English Speakers (100-Point Scale)

Rating	Salesperson		Cleaner		Server	
	Low	High	Low	High	Low	High
Competence	36.17 (28.63)	84.92 (27.62)	71.35 (32.02)	89.67 (20.98)	44.96 (32.44)	93.88 (6.90)
Treatment preference	18.17 (22.00)	90.67 (16.74)	44.09 (33.65)	94.63 (11.82)	38.91 (34.32)	92.75 (13.90)
Comprehensibility	94.08 (13.02)	96.42 (13.05)	98.50 (5.60)	99.29 (2.44)	98.38 (4.99)	98.21 (3.45)
Accentedness	98.41 (3.20)	98.65 (3.70)	97.50 (8.79)	99.54 (1.32)	94.87 (9.43)	98.13 (2.98)

Rating	Doctor		Lawyer		Professor	
	Low	High	Low	High	Low	High
Competence	61.58 (39.26)	92.71 (20.49)	67.09 (31.81)	93.75 (9.72)	50.13 (33.57)	95.63 (5.90)
Treatment preference	34.61 (38.98)	94.67 (11.04)	49.83 (34.97)	95.17 (7.55)	21.08 (29.95)	94.35 (8.60)
Comprehensibility	83.17 (30.57)	99.42 (2.30)	93.67 (10.14)	99.25 (2.56)	99.13 (2.01)	97.67 (6.42)
Accentedness	97.87 (4.43)	99.33 (1.52)	94.58 (14.18)	97.58 (5.82)	94.92 (13.23)	97.67 (5.40)

Means (Standard Deviations) for Ratings of L1 Tagalog Speakers (100-Point Scale)

Rating	Salesperson		Cleaner		Server	
	Low	High	Low	High	Low	High
Competence	52.79 (33.54)	86.70 (13.69)	67.63 (25.71)	84.88 (21.74)	60.35 (24.64)	87.38 (24.41)
Treatment preference	42.00 (35.84)	83.67 (21.54)	54.52 (25.94)	89.92 (11.38)	48.13 (32.75)	96.38 (6.92)
Comprehensibility	82.13 (18.08)	83.92 (15.79)	71.73 (22.93)	77.88 (20.23)	80.38 (19.42)	87.50 (16.28)
Accentedness	46.46 (21.22)	58.36 (23.56)	41.23 (24.02)	40.87 (20.37)	53.75 (24.63)	56.04 (22.52)

Rating	Doctor		Lawyer		Professor	
	Low	High	Low	High	Low	High
Competence	59.00 (30.57)	92.83 (10.97)	73.74 (26.57)	86.71 (21.33)	52.29 (26.58)	83.25 (29.17)

Treatment preference	31.70 (28.27)	93.33 (12.59)	61.38 (36.48)	93.22 (10.19)	23.14 (24.67)	93.58 (11.10)
Comprehensibility	69.09 (25.36)	85.96 (19.15)	77.88 (23.77)	85.58 (15.31)	74.17 (20.73)	74.96 (18.60)
Accentedness	49.52 (25.02)	46.22 (22.09)	57.52 (24.75)	67.50 (19.01)	45.14 (19.03)	40.71 (24.76)

Address for correspondence:

Mary Grantham O'Brien

School of Languages, Linguistics, Literatures and Cultures

University of Calgary

D412 Craigie Hall 2500 University Dr. NW, Calgary, AB, T2N 1N4, Canada

mgobrien@ucalgary.ca

Co-author information:

Pavel Trofimovich

Department of Education

Concordia University

1455 De Maisonneuve Blvd. W., Montreal, QC, H3G 1M8, Canada

pavel.trofimovich@concordia.ca

Cesar Teló

School of Languages, Linguistics, Literatures and Cultures

University of Calgary

2500 University Dr. NW, Calgary, AB, T2N 1N4, Canada

cesar.telo@ucalgary.ca