

“To my surprise, I don’t particularly like my own opinions” – exploring adaptations of the ‘open-guise’ technique to raise sociolinguistic language awareness.

Mats Deutschmann,
School of Humanities, Education and Social Sciences,
Örebro University
mats.deutschmann@oru.se

Anders Steinvall,
Department of Language Studies,
Umeå University.

anders.steinvall@umu.se

Abstract

The following study describes a data driven learning scenario aimed at raising sociolinguistic awareness of matters related to gender, language and stereotyping. The design is inspired by the matched-guise technique (MGT), a quantitative data driven experimental method that has been used extensively to investigate language attitudes. In the scenario, differences in respondents’ response patterns to two gender-manipulated versions (male-female vs female-male dyads) of the same dialogue were used as a starting point for awareness-raising activities aimed at highlighting how gender stereotypes may affect perceptions of a dialogue. The main focus of the article is a comparison of the learning outcomes of two variants of the setup: a traditional MGT setup, where the design and purpose of the experiment was kept secret until after the response phase, and a so-called open-guise design, where respondents were informed of the design and purpose of the experiment prior to the response phase. Preliminary results suggest that respondents adjust their assessments of a speaker depending on the guise, even when they know it is the same speaker they are listening to. Further, the open-guise design seemed to lead to greater pedagogic impact than the scenario based on the MGT design. Further studies are however needed to confirm these findings.

Introduction

According to Gilquin (2021, 231), data driven learning (DDL) is advantageous since learning can be inductive and implicit, and based on exposure to authentic data. To date, DDL has primarily been associated with the use of corpus data in the language classroom, a practice which began as early as the 1960s with the advent of corpus linguistics (McEnery and Wilson 1997). To date, such learning designs have mainly been focussed on features associated with linguistic competence, and tangible language aspects such as grammar and lexicon have been the objects of enquiry/learning (Boulton and Cobb 2017, 380). To our knowledge, DDL has not featured extensively as a method in the more challenging socio-pragmatic aspects of second language learning.

This article explores the pedagogic potential of the matched-guise technique (hereafter MGT), a quantitative data driven experimental method that has been used extensively to investigate language attitudes. Below, we will give a brief overview and some critical reflections of some of the models that we have developed over the past few years in the project Raising Awareness through Virtual Experiencing (RAVE – funded by the Swedish Research Council), where we have used MGT inspired designs to raise learners' sociolinguistic language awareness by exposing them to their own language attitudes and stereotypes. The main focus of the article, however, is a description and evaluation of recent adaptations of the models, where we use a so-called *open-guise* design (Soukup 2013). This adaptation fixes some of the ethical and methodological dilemmas of the previously developed MGT inspired approaches, but also comes with its own challenges and potential shortcomings.

Background

Our methods are inspired by the MGT (see Lambert et al. 1960), and later developments of this method, where the use of digital technology has opened up new possibilities (see Connor 2008, for example). The MGT is a sociolinguistic experimental design that was initially developed to measure attitudes towards a specific language, dialect, or accent. The framework has been described as “rigorous and elegant” (Garrett et al 2003, 57), and has since its introduction been used in a plethora of studies, particularly in the fields of sociolinguistics and social psychology.

In a MGT set-up, the same text, normally a spoken monologue, is produced in two or more variants, where the manipulated variable is the perceived identity of the speaker as manifested through language output – the language spoken or a social/regional accent, for example. In order to eliminate as many unwanted background variables as possible, bilingual/bidialectal actors have traditionally been used to produce the recordings. The text versions are then played to respondents who do not know the real purpose and design of the experiment. They are also unaware of the fact that it is the same actor that has produced both/all text versions under investigation. The respondents are then asked to rate (on a Likert scale) their impressions of the ‘different speakers’ on a number of personality characteristics, usually related to the dimensions of competence and warmth (cf. Cuddy et al. 2008). The reactions elicited by each of the linguistic guises are then compared. Since, arguably, the only variable that varies between the different recordings is the language/dialect/accent, differences in reaction are attributed to the respondents' attitudes towards the varieties spoken, and thus, in extension, towards the social groups with which these varieties are associated.

While the MGT and the closely related verbal guise design (where different actors are used, see Kircher 2015) have been used extensively to explore student and teacher attitudes towards different dialects, accents and languages in educational contexts (see for example Buckingham 2014; Carrie 2017; Kim 2021), there are to our knowledge no examples of instances where the method has been employed as a pedagogical tool to raise linguistic self-awareness of these issues. Given the growing field of unconscious/implicit bias training (see for example Project Implicit 2011; Sleek 2018), this is somewhat surprising. A MGT experiment is relatively easy to set up and can be conducted in the classroom in a matter of minutes (see Kircher 2015). The analysis of the results is straightforward, and gives an indication of language attitudes directly relevant for the participating group. These findings can then be used as a point of departure for discussions, self-reflections and other awareness raising activities. It is thus our belief that pedagogic adaptations of the MGT can make a significant contribution to raising sociolinguistic awareness in anti-bias training.

There have, however, been several points of critique raised about the MGT. As early as 1971, Lee criticized the method for being overly artificial and thereby of little relevance to real-life language situations. This point of critique was particularly relevant given the fact that the stimulus usually consisted of a read passage and thus totally lacked situational context (see Bradac et al. 2001). The artificial nature of the language stimulus in the traditional MGT design is a direct result of the method prioritising control of background variables that may influence impressions (such as voice quality) over authenticity. The fact that the covert nature of the method has been a prerequisite for the design has also meant the exclusion of more authentic-like language production, such as spontaneous speech or a dialogue: it is unlikely that respondents would conceive almost identical versions of longer strings of spontaneous speech produced by supposedly different speakers as believable.

One way around this dilemma is to split respondents into two randomized groups, where each group gets to hear only one of the language samples under investigation (see Stefanowitsch 2005). This modification opens up for more complex stimuli, including dialogue, where context such as speaker roles and purpose can be incorporated. This type of procedure, which is the one we also have adopted in many of our method designs (see Lindvall-Östling et al. 2019), however, excludes within-subject comparisons of responses. The pedagogic impact of being able to illustrate to an individual how his/her own specific impressions may be influenced by aspects such as accent is thereby lost. Furthermore, larger groups of respondents are needed to illustrate statistically reliable differences in response patterns.

Another dilemma with the MGT has been that, even when working with the same actress/actor, it is impossible to entirely control for all unwanted background variance such as speed of delivery, intonation, pitch and other aspects related to voice quality (Tsalikis et al. 1991). Moreover, in so-called verbal guise designs, when out of necessity multiple actors are used (when investigating gender or when there is no possibility of locating believable bilingual/bidialectal speakers, for example), this challenge becomes even more problematic. Here recent developments in technology have afforded entirely new possibilities, which we also have used in our designs. For example, digital cut-and-paste techniques of key phonetic signal markers have enabled the creation of digitally manipulated guises based on the same recordings (see Fridland et al. 2004; Campbell-Kibler 2008; Labov et al. 2011). Similarly, digital voice pitch and timbre manipulations have enabled the simulation of male or female versions of the same recording (Levon 2007; Denhag et al 2019).

Yet another point of critique of the method is that the rating scale format of the evaluations compels respondents to look for contrast where they might not normally note it, and thereby risks evoking stereotypic judgements that would not actually exist in an authentic

unconditioned situation (see Luhman 1990). In our more recent designs, we have thus preceded the rating evaluations with free text evaluations of the general impressions that respondents get from listening to the recordings (see below).

A final point of critique of the MGT that deserves attention concerns the ethics of the method. As mentioned earlier, the design presupposes secrecy as to the real purpose of the experiment. Not only does this create an ethical dilemma whereby researchers have to 'trick' respondents initially, even if this is partially repaired when the research design is revealed in the post-experiment debriefing. The initial secrecy of the design also creates a number of practical challenges: it is at times difficult to create a believable fake context/motivation for the experiment that does not arouse suspicion; it is difficult to assure that respondents have not suspected the real intentions of the design (such responses should be excluded from the analysis according to Kircher 2015); secrecy also means that a respondent group only can be used once; the secrecy aspect means that respondents should not discuss the learning experience with potential future respondents, peers in the year below, for example. This is of course difficult to ensure. Initial secrecy also means that informed consent can only be obtained after the event, which at times is ethically and practically problematic (see below).

In order to address some of the dilemmas listed above, Soukup (2013) has challenged the unquestioned premise of MGT studies, which holds "that informants are to be kept ignorant of the fact that they are hearing the same speaker(s) over again using different accents, varieties, or languages" (268). In her experiments, Soukup abandons the secrecy protocol and openly informs respondents of the design prior to listening to the text versions. According to Soukup (281), her informants had "no problem at all in making sense of the fact that they were hearing the same speakers twice, using different linguistic varieties", and the rating patterns mirrored findings from other traditional MGT experiments investigating the same language variants. Her results show that listeners can make sense of one and the same speaker putting on "different 'coats' of identity" (282), and successfully and honestly contrast personal impressions of social meaning in spite of the open design. In our latter designs we have trialled this open-guise approach, and this is the primary focus of this article.

Overview of our previous pedagogic activities based on MGT designs

The RAVE project, under which most of activities have been conducted, approaches the challenge of finding ways to increase sociolinguistic awareness of issues related to language and stereotyping among teachers trainees and other professional programs (sociology and psychology, for example), so that metalinguistic knowledge be translated into professional practice. We have conducted studies using MGT-inspired designs in various fields, including gender and personality psychology (Dennhag et al. 2019; Hakelind et al. 2020), gender and sociolinguistics (Lindvall-Östling et al., "An Exploratory Study," 2020), culturally gendered stereotypes (Deutschmann and Steinvall 2020), and stereotyping of accented students in English (Lindvall-Östling et al., "That's not Proper English!", 2020) and Swedish (Deutschmann et al. forthcoming), to mention a few.

The awareness-raising activities to date follow more or less the same procedural design. In short, the design consists of five phases: exposure to the case; response to the case; a debriefing session; a discussion seminar; and a short written-reflection (for a more detailed account, see Lindvall-Östling et al. 2019 and Deutschmann and Steinvall 2020). For the exposure phase students are divided into two groups, by the teacher or by a randomizer when

accessing the case. The exposure to the case and the ensuing response phase take place online where students listen to a dialogue and immediately after give their impressions on one of the characters. This data is summarized and analyzed, and then presented in a debriefing seminar, which is immediately followed by a discussion seminar in which students can discuss results, their impressions and the implications of these. Finally we ask them to write short reflective comments on a few questions we have prepared online. Although successful with regard to stimulating discussion and reflections (see, for instance, Hakelind et al. 2020), the design has not been without issues of different character, which is one reason for testing the open guise method for pedagogical purposes.

One issue has been the secrecy. In order to create maximum effect at the debriefing/discussion seminar – an aha-moment – keeping students unaware of the real purpose of the task has been important: Accordingly, the activity has been contextualized in relation to a particular theme of the course or context which is different from the main focus of the activity. For example, in Lindvall-Östling et al., “That’s not Proper English!”, (2020) the activity was framed as an evaluation exercise of oral performance, while the actual purpose was to find out whether different national accents (British RP vs Indian accent) affected the judgement of personality characteristics. The design of the awareness-raising activities has thus to some extent been dictated by the ambition to keep the true purpose hidden from the students.

Another pedagogical issue has been the fact that in any class half of the students listened to one configuration of the guises and the other half listened to the opposite configuration. Although pairing students for the debriefing discussions of impressions and results have been effective, each individual could only truly relate to the version they themselves had listened to. In exposing students to both versions, they would have the opportunity to reflect on their own reactions, especially if the entire setup was clear to them.

From a practical point of view of research, informed consent can only be obtained after the debriefing, of course, when students are aware of the full picture. When this does not happen directly after the response phase, the data from respondents that are absent at the debriefing or who do not answer the post-survey cannot be used.

Aims

The main ambition of this article is to describe recent pilot adaptations of the MGT-inspired framework summarised above to a so-called open-guise technique (Soukup, 2013), where the respondents are informed of the design and purpose prior to the experiment. We will also compare evaluations of the learning experiences of two groups of students who participated in similar scenario set-ups using a MGT-design and an open-guise design.

Description of activities

Script and recordings

The setup described here was initially developed for a program in language, rhetoric and communicative leadership, but has also been piloted in courses in sociolinguistics in teacher training programs. The focus of the exercise is on communication and leadership, and how gender stereotypes may influence our impressions of a communicative event. The speech sample is inspired by dialogues described in Holmes (2005). It consists of a workplace

interchange where a boss (Robin) tells an employee (Kim) off for not doing his/her job properly:

Robin: I assume this sort of stuff is backed up on the secure internal server, right?

Kim: Eerm. I'm.. I'm not sure.

Robin: What do you mean "you're not sure"?!

Kim: Well, eerm, I mean John and Beth are the ones that are involved with security and back-ups so ...

Robin: So if they weren't here we'd be totally lost, right... and you wouldn't have a clue!?

Kim: I'd most probably look up the formal internal routines for this sort of thing... that don't exist...

Robin: Well... Jesus! You're telling me you don't know, or worse, that there are no routines – this is a critical issue, don't you think? If we lose this type of stuff, or, just imagine if it ends up in the wrong hands! We are talking major disaster! Things can't be run like this!

Kim: No, I guess not. Sorry, I'll try to look into it.

Robin: Don't try Kim! Just do it! Give me an overview of the routines when you're done.

The script was recorded on separate tracks using a female actress enacting both the characters Robin and Kim. Initial audio editing was made using the software *Twisted Wave* (<https://twistedwave.com>) to remove unwanted pauses and sounds e.g coughing or pauses and/or repetitions made due to reading mistakes etc. The voice quality of the recordings were then manipulated using the software *Melodyne*¹, a professional post-editing tool used in the music industry. In *Melodyne* the audio is represented diagrammatically, which enhances the ability to tweak the audio in multiple ways. Consequently it is possible to vary aspects such as the pitch, vibrato, volume, sibilant length, timing and formants of each single note. The main tool used within *Melodyne* was Pitch transition. Initially, the whole monologue was pitch transitioned (this is popularly known as *auto tuning*). Manual pitching was then used to adjust parts that sounded unnatural in an attempt to minimize the risk that listeners perceived the audio as manipulated. It is, however, near impossible to produce a perfectly natural sounding manipulated voice. At the end of manipulation, we had two different feminine sounding voices and two different masculine sounding voices, all of which had been manipulated. These were combined in different ways producing a total of four set-ups: M-F (Robin-male: Kim-female); F-M; M-M; F-F².

The recordings were 'packaged' in youtube videos depicting male/female silhouette figures engaged in a conversation in an office environment (see Figure 1 below).

¹ <https://www.celemony.com/en/melodyne/what-is-melodyne>

² The recordings are available on our website: <https://www.stereotyping.se/>

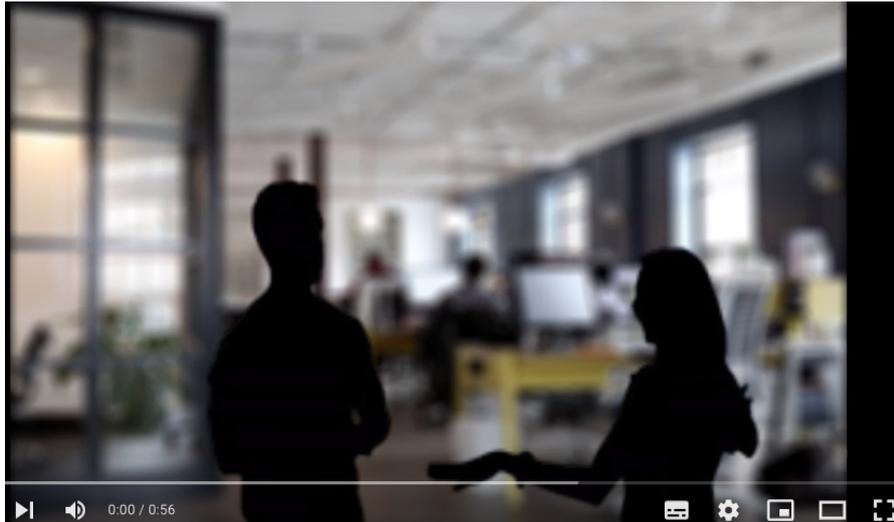


Figure 1. Example of image background to recordings (F-M version).

The response questionnaire

After the production phase described above, the videos were embedded in online questionnaires (SurveyMonkey), where they constituted the response stimuli. In the questionnaires, participants are first asked to give their spontaneous free text reactions to the dialogue with focus on the communicative styles and characters of Robin and Kim.

This is followed by a set of statements exploring different characteristics of Robin (the leader), and/or Kim (the employee). Respondents are asked to agree or disagree with these statements on a 7-point Likert scale, where 1 represents complete disagreement and 7 complete agreement. They are formulated as positive statements according to the format *Robin/Kim is + adjective or descriptive phrase*, and statements in the trials included descriptions relating to leadership and communicative aspects positioned on the dimensions of competence (for example, *Robin is ...effective, a good leader, straightforward and clear etc*) and sympathy/warmth (for example, *friendly, abusive, rude* - see Figures 4 and 5 below).

In the MGT set-up several versions are included in the questionnaire (M-F and F-M, in this case) and a randomising tool decides which version respondents get to listen to. In the open-guise setup we instead produce separate questionnaires for each version. The participants are also given the opportunity to comment on any aspects of the exercise, and in the open guise version they are also asked to give consent (or not) that their responses be added to the research database. This is unfortunately not possible in the MGT design, as respondents are unaware of the real purpose of the exercise. Informed consent thus has to be postponed to the post-debriefing survey (something which has also led to considerable loss of data - see below).

For practical reasons, we have so far only trialled comparative scenarios exploring differences in impressions of mixed-sex setups (F-M vs M-F). In addition, the primary focus of the quantitative analysis (statement responses) so far has been on the leader (Robin). Of course various other comparative combinations are possible, including a comparison of all four versions with focus on both Robin and Kim. Based on previous experience from the project (see Lindvall-Östling et al. 2019), however, we have found that trying to include too many aspects in one scenario lessens the pedagogic impact of the exercise and causes confusion.

Pilot trials

The descriptions and data below are based on two pilot runs of the scenario that we did with classes of student teacher trainees studying a course in sociolinguistics in autumn 2020 and autumn 2021. One group (N=30) did the exercise as a traditional MGT setup. They were thus not informed of the design prior to the experiment. This group were randomly assigned to either the M-F version or the F-M version of the recording. In the second trial (N=19), we used an open-guise design and told the participants about the purpose of the exercise prior to the experiment. Participants were given access to both versions (M-F and F-M in a counterbalanced design), but were told to wait at least a day between the listening occasions in order to minimise interference of the previous impressions. In the open-guise trials, participants were also asked to give informed consent that their responses be added to our research database at this stage.

In both pilot runs, the debriefing seminars, where we revealed the design (relevant for the MGT trial only) and presented the response patterns, took place in the online video conferencing tool *Zoom*. The discussion seminar immediately following this took place in so-called breakout rooms (3-4 participants in each group) followed by a whole class discussion. Finally, participants were asked to give their reflections of the learning experience in a post-survey. It was also here that the MGT pilot participants were asked to give their informed consent that the data generated by the entire trial be added to the research database. Unfortunately relatively few respondents from the MGT trial (14 of 30) answered the post-survey. In the open-guise trial 15 of 19 participants completed the post-survey.

Debriefing material based on student responses to the recordings

Below follows examples of the type of material that was presented to the participants in the debriefings of the pilot trials of this case scenario. The material is taken from the response patterns generated in the open-guise pilot trial. As such, it constitutes a summary of the group responses to the recordings, and aims to illustrate how responses systematically differed (or not) depending on the version of the recording. Note that the debriefing material from the MGT trial is not presented here since consent to publish this material is missing from 16 of the 30 participants. The response patterns from the MGT trial were, however, very similar in character to those generated in the open-guise trial. Finally, given the fact that the focus of this article is on pedagogic design and the students' learning experience, we have not included any detailed quantitative statistical comparative analysis of the responses of the two trials.

Firstly, the qualitative descriptions in the questionnaires were summarised and visualized using a word cloud tool (worditout.com). This was done by extracting all the descriptive phrases and adjectives that occurred in the free text descriptions of the characters Robin and Kim (N=19) for each guise version and saving these in separate text files for each version of the recording (M-F and F-M). These text files were subsequently used to generate word clouds (see figures 2 and 3 below). The word cloud software creates an image, where the size of the words are indicative of how frequently they occur in the text.

In both pilot runs, it was evident that the descriptions of male Robin were more negative and emotionally oriented than those of female Robin (see Figure 2). For example, Robin was referred to as *rude* five times in the male description, as *condescending* three times and as a *bully* three times in descriptions of the male guise. In contrast, adjectives, such as *concerned* (five times), *upset* (five times) and *tired* (5 times), that somehow justified Robin's behaviour as an understandable reaction to Kim's incompetence were frequent in the descriptions of the female guise (see Figure 2 below)

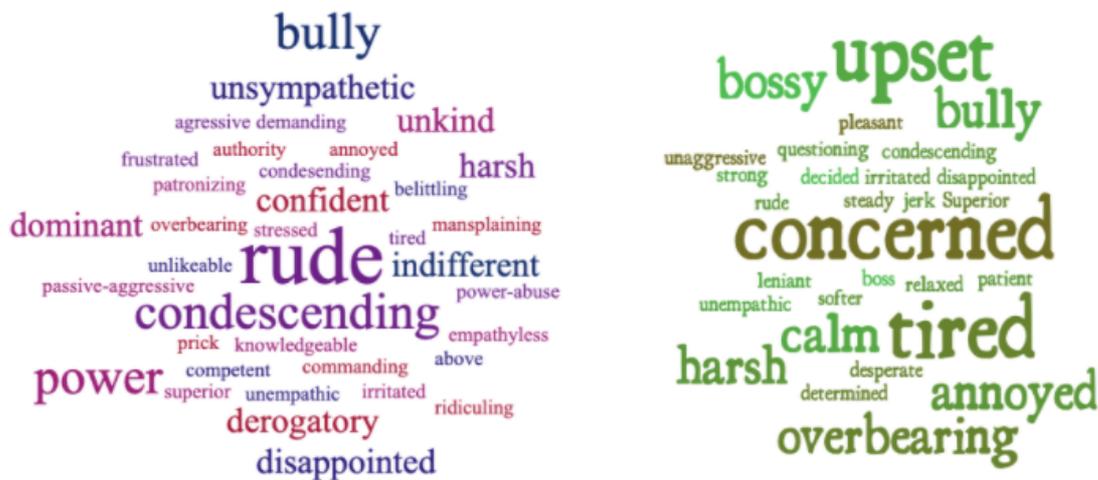


Figure 2. Word clouds of adjectives and descriptive phrases used to describe Robin (the boss) after listening to the male guise (purple-blue) and the female guise (green)

There were clear differences in the descriptions of the female and male versions of Kim too. Female Kim descriptions were dominated by adjectives such as *insecure* (5), *nervous* (3) and *submissive* (2) (see Figure 3 below). This semantic field also occurred in the descriptions of male Kim, but here negative descriptions of Kim’s professional competence (*lazy*, *careless* and *blasé*) were also frequent. Also worth noting is that the most frequent adjective used to describe male Kim was *weak* (5 occurrences).



Figure 3. Word clouds of adjectives and descriptive phrases used to describe Kim (the subordinate) after listening to the female guise (purple-blue) and the male guise (green)

The quantitative data based on the statement responses mirrored the findings from free text data above. In the responses to the male guise, there were stronger tendencies to agree with statements referring to negative characteristics such as being *rude*, *insensitive*, *arrogant*, *aggressive*, or a *bully* when the respondents listened to the male version. In contrast, the female version was ranked more positively with reference to professional aspects such as *competence*, being a *good leader* and being *effective* (see Figure 4 below).

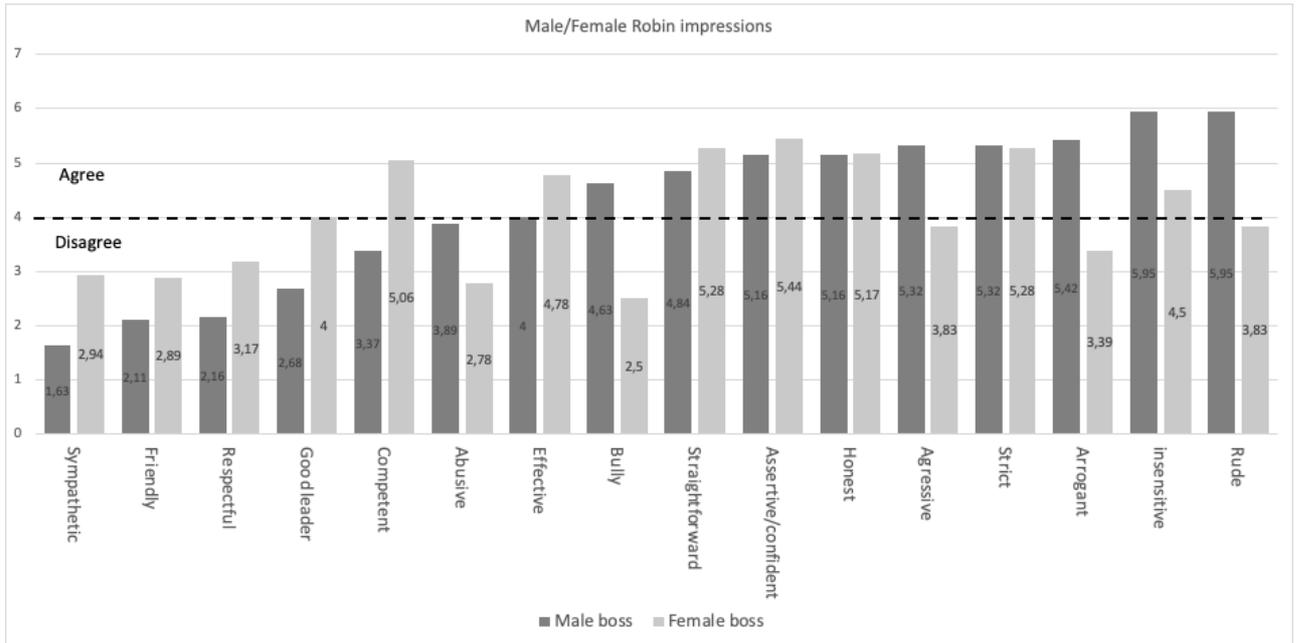


Figure 4. Impressions of traits of 'male Robin' (dark bars) and 'female Robin' (light bars). Note that 4 represents a neutral (neither nor) alternative on the Likert scale.

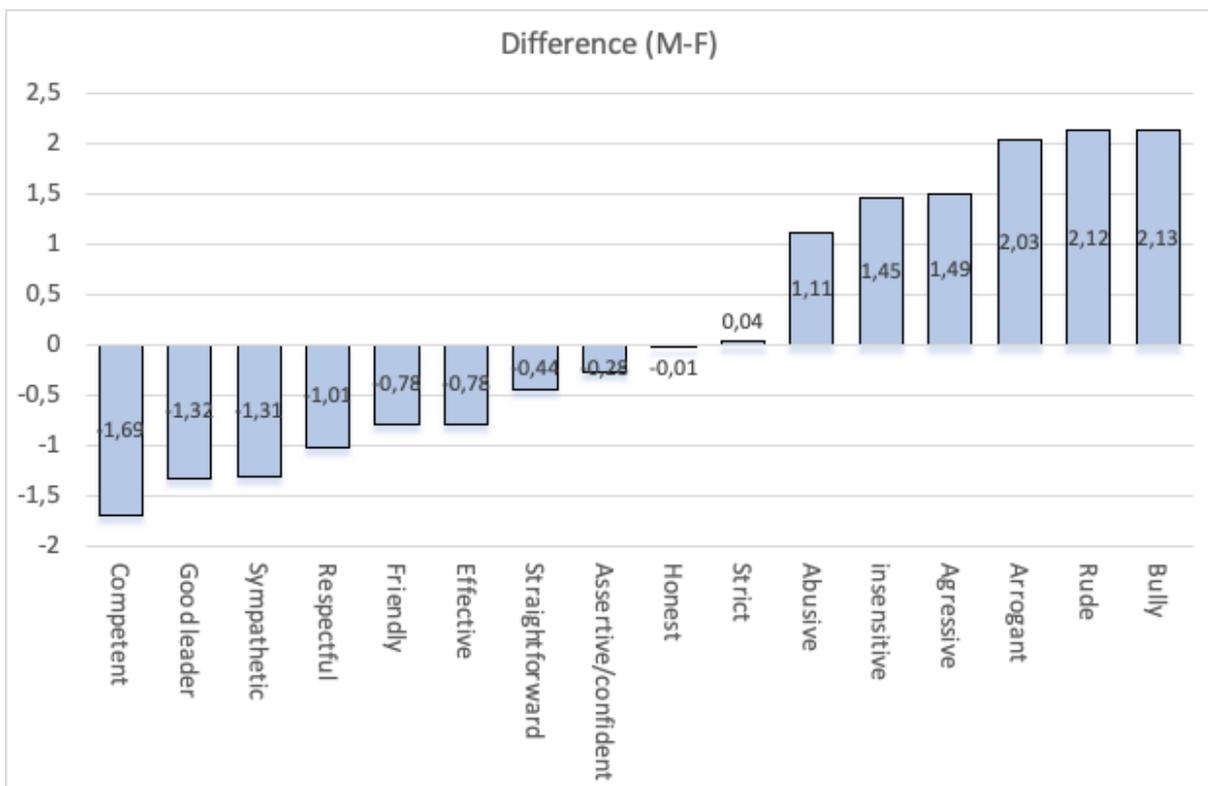


Figure 5. Differences between 'male Robin' and 'female Robin' ratings (M-F) in order of magnitude starting with negative values where 'female Robin' was rated more favourably than 'male Robin'

The largest differences in ratings between the male and female versions of Robin were observed for the traits *competence*, *being a good leader* and *being sympathetic* where the female version received more favourable ratings. For the traits *arrogance*, *rudeness* and *being a bully* the male version received higher ratings. Figure 5 above summarises the mean differences (male values - female values) between ratings of the male and female versions of the recording. Negative values imply that the male version was rated lower on the traits and positive values that it was rated higher.

Evaluating the learning experiences - MGT and Open-guise compared

The results below are based on the evaluations of 14 respondents who partook in the MGT trial and 15 respondents who partook in the open-guise trial. The results below include a summary of responses to the following questions in the post-survey aimed at capturing participants' impressions and reflections of the learning experience:

1. *Did the activity you just participated in give you any new insights? If so, what were they?*
2. *In what way can the experiences you gained from this experiment help you in your (future) profession?*
3. *Was there anything in the design of this activity that you feel worked particularly well, or alternatively that worked less well and you feel that we should change?*

In addition, the survey included a "General Comment" where respondents could leave any other reflections or comments.

Question 1 - New Insights

The most common answer (8/14) in the MGT group was that the exercise had not led to any new insights. Many of these responses were simply formulated as short negative answers (*No*; *No, not really*; *Not really*), while others provided more comprehensive explanations that pointed to the fact that they already knew about these stereotypes but that a reminder was good, for example:

- *Nothing inherently new. This mostly proves the many hypotheses and theories that we've learned from before;*
- *Not really, I've done this types of tests before but it's always good to freshen up on your thinking about language;*
- *In this particular case, no. I'm already aware of (most of?) my preconceptions in regards to gender.*

Six respondents in the MGT group did however refer to new insights, and in all these cases these were self-reflections related to how the masculine version was evaluated more negatively. For example:

- *I have come to realise that I think it's worse when a man argues with a woman instead of the opposite;*
- *I focused more on power than gender, but I would still say that I have stereotypical preconceptions concerning gender, and I realise that I interpret some behaviors more or less harsh depending on whether it is a man or a woman;*
- *When trying to be equal towards women the scale might fall to the other side, we might start treating women better than men and not equal.*

In the open-guise there were 4/15 negative answers. Only one of these responses was a short answer (*No*). The remaining three responses referred to the importance of this type of reminder, for example:

- *Mostly, I feel like it worked as a reminder of something I already knew, but often forget/don't think about - further proving that this is something worth being reminded of (that we're influenced by stereotypical preconceptions);*
- *They rather confirmed an existent assumption.*

The remaining responses (11/15) mirrored those of the MGT groups and were generally examples of self-reflections. There were also several responses expressing surprise, for example:

- *We are more judgemental than I thought we were!;*
- *I realise that my expectations and presumptions about gender connected stereotypes are certainly not stable and correct;*
- *Yes, that I am quite judgemental and have fixed stereotypes. I did think more of myself, but I noticed that I really was judgemental!;*
- *"I've answered as truthfully as I can in this survey, and to my surprise I don't particularly like my own opinions in a few of these cases. I thought I was less biased in regards of gender stereotypes as I'm against them in general.*

In summary, the open-guise evaluations were more positive than the MGT evaluations for the first question. For example, expressions that pointed to new insights were more common in the open-guise trial. It was also among these respondents that we saw frequent expressions of surprise at the insights gained.

Question 2 - Relevance of learning experience to future professional practice

Many answers to this question in the MGT group (4/14) addressed aspects related to how the students as future teachers would use the knowledge to monitor their own behaviour in order to avoid bias, for example:

- *I will think about my perceptions about people and language a bit more and try to have an open mind;*
- *It has taught me that you should always question why you do a certain thing, is it equal for everyone?;*
- *It has taught me that it is important to treat students, whatever the gender, with equal respect and pedagogical approach.*

There were also a number of answers that concerned more general aspects of sociolinguistic awareness, for example:

- *It has made me more aware of how communication between the sexes is perceived. What might be considered okay for a woman to say to a man might not be okay for a man to say to a woman, for example;*
- *It has taught me to see all people for who they are, not their gender;*
- *It has shown how much the style of communication actually matters in not only my profession but overall. How we are perceived by others, and how relationships can be affected by this.*

Finally, there were two examples of respondents saying that they would use a similar setup in their own future teaching:

- *I might actually use something very similar when speaking about social constructs (concerning gender, race, or other) in upper secondary school;*
- *Perhaps if I am working with norms I can use a similar exercise with the students.*

Just as in the previous group, there were several examples of how raised awareness would be applied in future professional contexts in the open-guise group:

- *As a teacher student this experiment helped me see what preconceptions I have of others and also think of the preconceptions others can have of me. In the future as a teacher I will have to be conscious about the judgment I have of my students and coworkers while also considering the expectations both students and coworkers will have of me;*
- *I might not "hyper-correct" how I let speakers take space in my classroom (e.g. letting boys speak too little because I expect them to speak too much etc);*
- *I think it is an important topic to talk about and be aware about. It is good to have an understanding/knowledge about how students might perceive you as a leader but also how you see your students when it comes to gender;*
- *This has opened my eyes to stereotypes and how we should work with them in the classroom.*

There were also a few examples of more general insights gained from the activity that did not necessarily directly relate to teaching:

- *That I should not be that judgemental, and that I should be more objective and take the situation for what it is;*
- *It has taught me to reflect more about these questions;*
- *Personally, the experiences gained today will hopefully serve as a reminder to myself to be more aware of my own prejudices and stereotypes. And for me to check the way I speak or the way I judge others depending on their speech more regularly.*

Finally, there were five examples of students expressing intentions to use similar set-ups in their teaching in the open-guise group:

- *This experience has given me ideas on how to, in an effective way, show students that stereotypes and gender norms apply to them as well;*
- *I will probably do a similar experiment in my classroom in order to create awareness of their own expectations on gender related to professionalism;*
- *I want to do this kind of exercise with my students! I feel like it would be a fairly simple and effective way of making students more aware of sociolinguistics and how men and women are perceived. Especially since the subject is something that, in my experience, teachers sometimes struggle to bring into their course plans, this could be an intriguing thing to do in the classroom!*
- *I could implement this "raising awareness about stereotypes" for the "värdegrundsarbete" to make students understand what stereotypes could cause and also the benefit of reducing them.*
- *As mentioned, we often tend to think about "talking about norms/stereotypes" as central but we have to develop both the education for us as teachers and when we start working to create tools in how we behave and make decisions. We must work more actively with these types of "challenges" in an everyday classroom. I truly believe this example shows how we can work with these important issues more concretely and this should be more central when educating future teachers!*

In summary, both groups could exemplify how the exercise would have a direct impact on their future professional approaches. The open-guise group were arguably slightly more enthusiastic over the method as illustrated by the fact that 5/15 respondents actually said that they would like to do something similar in their own teaching.

Question 3 - Methodological reflections

Six respondents in the MGT-group mentioned the quality of the voices in the recordings as an issue that may have influenced them in their impressions of the recording, for example:

- *I think the discussion worked well regardless of being on Zoom. I think I noticed some issues with the voice manipulation, and maybe that affected me. You could opt for having two actors, one male and one female, but then again you would not have the same exact delivery in the lines of each actor;*
- *I could tell that Robin's voice had probably been altered in some way, which made it hard for me to focus on my initial impression of the voice. That's probably not ideal;*
- *The male voice of Kim felt fake, which made it less of a real experience than it could've been;*
- *The quality of the voices could be better.*

Apart from these negative comments related to voice quality, there were other points of critique raised. For example, some thought that the recordings were too short, while one respondent thought that it would have been better to use 'authentic dialogue'. Another respondent made the point that we may have missed important aspects of the impressions since we dictated the focus through the choice of statements (but note that we did invite free text answers too). These points are illustrated in the extracts below:

- *I would have liked for the video to be a bit longer so that you had a better picture of the people;*
- *Is there a way of using "real" dialogues? Would make it feel more authentic, may make the result more trustworthy (some may argue the current version feels "forced");*
- *While it is good to "force" a choice in grading type questions, be aware that many contextual things may fly under the radar.*

Interestingly, one respondent also questioned the between-subject comparative design and suggested that we develop an open-guise version of the experiment:

- *How accurate are the results since we only get to listen to the conversation between either a man and a woman or a woman and a man? Could it be an idea to get everyone to listen to both recordings? I mean, since everyone has different thoughts, opinions etc.*

Finally note that there were several positive comments too:

- *I think it was very interesting and worked well;*
- *The diagram was very interesting to analyze.*

Overall, the comments on methods from the open-guise group were more positive. There was only one respondent that commented on the quality of the voices being an issue:

- *The manipulated voice sounded too manipulated and was hard to sympathize with.*

Instead some respondents specifically noted that knowing that the voices were manipulated actually made them think:

- *I liked that I knew beforehand that the recordings were the same person and it was just manipulated. It made me think about how I perceive male and female voices.*
- *I did also think it was interesting to do the experiment while knowing that I would react to male-female and female-male interactions. I had to examine and reflect over my judgements and expectations based on the genders of the speakers.*

Other self-reflective positive comments included curiosity of other potential outcomes with other gender combinations or research designs:

- *No, I think everything was perfect! But it would have been fun to see how my impression would change if I heard a recording of Kim AND Robin being male, and vice versa. Would Robin (M) be seen as aggressive if Kim also was a man? etc;*
- *It would be interesting to see if a group that was unaware of the manipulations had similar opinions.*

The open-guise groups' comments in this category also included many general positive statements commenting on different aspects of the design, for example:

- *I think that the study I took part in made me understand my expectations about how different genders act. The discussions in break-out rooms were interesting in the sense that our impressions were different;*
- *I liked the concept about this whole activity; the survey did not take up much time before the seminar which I think is important so that you feel that you have the interest to answer thoughtfully. I also liked the concept of the seminar, easy to follow and good to have smaller groups to discuss and then lift the important points together.*

In summary, the responses from the open-guise groups were more positive and there were less indications that the voice quality of the recordings was perceived as a problem (of course both groups listened to the same recordings).

The final comments from both groups included several short positive statements such as:

- *A good exercise. An eye-opener!*
- *Interesting topic to discuss, good teacher.*
- *Thanks for the interesting lab!*
- *Overall, a pretty fun exercise. Well done!*
- *very interesting lab.*

Overall summary

The evaluations from the exercise show that although a fair proportion of the students did not see how the exercise led to any new insights (especially in the MGT group), the vast majority, in both groups, could see how the knowledge gained was directly applicable to their future roles as teachers. Some students, especially in the open-guise group, were also inspired by the design and said they wanted to use it themselves in the future. Regarding the reflections on method, many students in the MGT group especially, felt that the voice quality of the manipulated voices was problematic. This critique was less obvious among the open-guise group.

Discussion

Before moving on to discuss the merits/drawbacks of the open-guise method based on our findings, it is important to point out some serious limitations in the current study. The findings so far are based on very limited data from two trials of the case scenario described above. We need to conduct many more trials in order to establish reliable data for response patterns so that we can elucidate what potential effects the open-guise vs MGT designs may have on perception and on the learning experience. It may be the case, for example, that expectancy factors (Feingold 1994), influence respondents to a greater degree in the open-guise design

than the MGT design. Expectancy factors, i.e. social and cultural expectations of behaviours, may influence respondents to respond 'appropriately', rather than honestly, especially when they know that the focus of the exercise is on gender and communicative styles. If this is the case, the main purpose of the exercise, i.e. to reveal hidden stereotypes within a group, is lost. Further, the limited data does not allow us to fully explore how the order of exposure (M-F or F-M) affects response patterns. Similarly, much more data is needed to confirm similarities/differences in the evaluations of the learning experience, and to capture all potential aspects of this experience; drawing conclusions on the qualitative responses of 14 or 15 participants is precarious to say the least.

Notwithstanding these limitations the results are interesting. Our response data corroborate Soukup's findings that the "open-guise technique actually 'works'" (Soukup 2013, 281): just as Soukup concludes, we can confirm that respondents adjust their assessments of a speaker depending on the guise, even when they know it is the same speaker they are listening to. The question why they do so, however, remains, and based on the limited data we have we cannot dismiss that expectancy factors may influence respondents.

The main focus of this study was however not on the differences of perceptions of the recordings per se, but rather on how well the open-guise vs MGT designs work as pedagogical methods to raise sociolinguistic self-awareness. Here our results suggest that awareness of the design and the exposure to both versions increased the impact of the learning experience. We can speculate as to why this should be the case.

Firstly, it is reasonable to assume that the open-guise design creates a greater potential for self-reflection, given the fact that respondents can compare how they themselves felt about the two versions and also relate this to the response patterns of the group as a whole. As pointed out earlier, one of the disadvantages of the MGT-design we have used to date is that respondents only listen to one version of the guise making within-subject comparisons impossible.

Secondly, it seems that voice quality issues become less of an inhibitory factor when respondents know that the voice is manipulated. This is reasonable. Not being told about the voice manipulation prior to the experiment may well lead to respondents seeing this as a primary cause for differences in interpretations (rather than gender stereotyping effects). After all, no one likes to be 'tricked' into exposing weak spots, and when this is the case we may subconsciously try to find prevarications to explain our behaviour.

Finally, we would argue that the logic behind the exercise was easier to convey in the open-guise design, thereby better preparing the students to fully understand the data presented to them and the logic of the exercise. Although not evident in this particular study, our experience from previous RAVE activities is that it is sometimes difficult to bring home to students the fact that they have in fact been exposed to different versions of the SAME recording.

In summary, we can conclude that the results from preliminary pilot trials discussed above are promising. The open-guise method unclutters some of the ethical and methodological dilemmas of the MGT design and allows us to more openly explore the multi-dimensionality of stereotyping effects. Further trials are however needed to assure the quality of the method.

List of references

- Boulton, Alex, and Tom Cobb. 2017. "Corpus Use in Language Learning: A Meta-Analysis." *Language Learning* 67 (2), 348–93. <https://doi.org/10.1111/lang.12224>
- Bradac, James J., Aaron Castelan Cargile, and Jennifer S. Hallett. 2001. "Language Attitudes: Retrospect, Conspect and Prospect." In *The New Handbook of Language and Social Psychology*, edited by W. Peter Robinson and Howard Giles, 137–55. New York, NY: John Wiley and Sons.
- Buckingham, Louisa. 2014. "Attitudes to English Teachers' Accents in the Gulf." *International Journal of Applied Linguistics* 24 (1): 50–73. <https://doi.org/10.1111/ijal.12058>
- Campbell-Kibler, Kathryn. 2008. "I'll Be the Judge of That: Diversity in Social Perceptions of (ING)." *Language in Society* 37 (5): 637–59. <https://doi.org/10.1017/S0047404508080974>
- Carrie, Erin. 2017. "‘British is Professional, American is Urban’: Attitudes towards English Reference Accents in Spain." *International Journal of Applied Linguistics* 27 (2): 427–47. <https://doi.org/10.1111/ijal.12139>
- Connor, Robert T. 2008. "Computer Simulation and New Ways of Creating Matched-Guise Techniques." *International Multilingual Research Journal* 2 (1-2): 102–08. <https://doi.org/10.1080/19313150802010319>
- Cuddy, Amy J.C., Susan T. Fiske, and Peter Glick. 2008. "Warmth and Competence as Universal Dimensions of Social Perception: The Stereotype Content Model and the BIAS Map." *Advances in Experimental Social Psychology* 40: 61–149. [https://doi.org/10.1016/S0065-2601\(07\)00002-0](https://doi.org/10.1016/S0065-2601(07)00002-0)
- Dennhag, Inga, Anders Steinvall, Camilla Hakelind, and Mats Deutschmann. 2019. "Exploring Gender Stereotypes about Interpersonal Behavior and Personality Factors Using Digital Matched-Guise Techniques." *Social Behavior and Personality* 47 (8): 1–13. <https://doi.org/10.2224/sbp.8150>
- Deutschmann, Mats and Anders Steinvall. 2020. "Combatting Linguistic Stereotyping and Prejudice by Evoking Stereotypes." *Open Linguistics* 6 (1): 651–71. <https://doi.org/10.1515/opli-2020-0036>
- Deutschmann, Mats, Eric Borgström, Daroon Yassin Falk, Anders Steinvall, and Johan Svensson. Forthcoming. "It Ain't What you Say it's the Way that you Say it – Using Matched Guise Technique (MGT) to Raise Awareness of Accent Stereotyping Effects among Swedish Pre-Service Teachers." *Language Awareness*.
- Feingold, Alan. 1994. "Gender Differences in Personality: A Meta-Analysis." *Psychological Bulletin* 116 (3): 429–56. <https://doi.org/10.1037/0033-2909.116.3.429>

Fridland, Valerie, Kathryn Bartlett, and Roger Kreuz. 2004. "Do you Hear What I Hear? Experimental Measurement of the Perceptual Salience of Acoustically Manipulated Vowel Variants by Southern Speakers in Memphis, TN." *Language Variation and Change* 16 (1): 1–16. <https://doi.org/10.1017/S0954394504161012>

Garrett, Peter, Nikolas Coupland, and Angie Williams. 2003. *Investigating Language Attitudes: Social Meanings of Dialect, Ethnicity and Performance*. Cardiff: University of Wales Press.

Gilquin, Gaëtanelle. 2021. "Using Corpora to Foster L2 Construction Learning: A Data-driven Learning Experiment." *International Journal of Applied Linguistics* 31 (2): 229–47. <https://doi.org/10.1111/ijal.12317>

Hakelind, Camilla, Anders Steinvall, and Mats Deutschmann. 2020. "The Power of Aha! On Stimulating and Guiding Students towards Self-Awareness and Critical Reflection while Teaching about Personality Psychology and Gender Stereotypes." *Psychology Learning & Teaching*: 1–16. <https://doi.org/10.1177/1475725720979460>

Holmes, Janet. 2005. "Leadership Talk: How do Leaders "Do Mentoring", and Is Gender Relevant?" *Journal of Pragmatics* 37 (11): 1779–800. <https://doi.org/10.1016/j.pragma.2005.02.013>

Kim, Sugene. 2021. "English as a Lingua Franca in Japan: Multilingual Postgraduate Students' Attitudes towards English Accents." *Journal of Multilingual and Multicultural Development*. <https://doi.org/10.1080/01434632.2021.1909053>

Kircher, Ruth. 2015. "The Matched-Guise Technique." In *Research Methods in Intercultural Communication* edited by Zhu Hua, 196–211). John Wiley & Sons, Inc. <https://doi.org/10.1002/9781119166283.ch13>

Labov, William, Sharon Ash, Maya Ravindranath, Tracey Weldon, Maciej Baranowski, and Naomi Nagy. 2011. "Properties of the Sociolinguistic Monitor." *Journal of Sociolinguistics*, 15 (4): 431–63. <https://doi.org/10.1111/j.1467-9841.2011.00504.x>

Lambert, W.E., R.C. Hodgson, R.C. Gardner, and S. Fillenbaum S. 1960. "Evaluational Reactions to Spoken Language." *Journal of Abnormal and Social Psychology* 60 (1): 44–51.

Lee, Richard R. 1971. "Dialect Perception: A Critical Review and Re-Evaluation." *Quarterly Journal of Speech* 57 (4): 410–7. <https://doi.org/10.1080/00335637109383086>

Levon, Erez. 2007. "Sexuality in Context: Variation and the Sociolinguistic Perception of Identity." *Language in Society* 36 (4): 533–54. <https://doi.org/10.1017/S0047404507070431>

Lindvall-Östling, Mattias, Mats Deutschmann, and Anders Steinvall. 2019. "'Oh It Was a Woman! Had I Known I Would Have Reacted Otherwise!': Developing Digital Methods to Switch Identity-Related Properties in Order to Reveal Linguistic Stereotyping." In *Virtual Sites as Learning Spaces: Critical Issues on Linguaging Research in Changing Eduscapes*,

edited by Sangeeta Bagga-Gupta, Giulia Messina Dahlberg and Ylva Lindberg, 207–39. England: Palgrave Macmillan. https://doi.org/10.1007/978-3-030-26929-6_8

Lindvall-Östling, Mattias, Mats Deutschmann, and Anders Steinvall. 2020. “An Exploratory Study on Linguistic Gender Stereotypes and their Effects on Perception.” *Open Linguistics* 6 (1): 567–83. <https://doi.org/10.1515/opli-2020-0033>

Lindvall-Östling, Mattias, Mats Deutschmann, Anders Steinvall and Satish Patel. 2020. “‘That’s not Proper English!’ : Using Cross-cultural Matched-guise Experiments to Raise Teacher/Teacher-trainees’ Awareness of Attitudes Surrounding Inner and Outer Circle English Accents.” *Educare – Vetenskapliga Skrifter* 3: 109–141. <https://doi.org/10.24834/educare.2020.3.4>

Luhman, Reid. 1990. “Appalachian English Stereotypes: Language Attitudes in Kentucky.” *Language in Society* 19 (3): 331–48. <http://www.jstor.org/stable/4168150>

McEney, Tony, and Andrew Wilson. 1997. “Teaching and Language Corpora (TALC).” *ReCALL* 9 (1): 5–14.

Project Implicit. 2011. Project Implicit. <https://implicit.harvard.edu/implicit/> (accessed October 20, 2021).

Sleek, Scott. 2018. “The Bias Beneath: Two Decades of Measuring Implicit Associations.” *APS Observer* 31 (2).

Stefanowitsch, Anatol. 2005. “Empirical Methods in Linguistics: The Matched Guise Technique.” http://socialperspectives.pbworks.com/f/exp_matchedguise.pdf (accessed October 12th, 2021)

Soukup, Barbara. 2013. “On Matching Speaker (Dis)guises – Revisiting a Methodological Tradition. In *Language (De)standardisation in Late Modern Europe: Experimental Studies* edited by Tore Kristiansen and Stefan Grondelaers, 267–85. Oslo: Novus.

Tsalikis, John, Oscar W. deShields, and Michael S. LaTour. 1991. “The Role of Accent on the Credibility and Effectiveness of the Salesperson.” *Journal of Personal Selling & Sales Management* 11 (1): 31–41.