Journal of Language and Social Psychology
http://jls.sagepub.com/

Reverse Linguistic Stereotyping: Measuring the Effect of Listener Expectations on Speech Evaluation

Okim Kang and Donald L. Rubin

Journal of Language and Social Psychology 2009 28: 441 originally published online 25 September 2009
DOI: 10.1177/0261927X09341950

The online version of this article can be found at:
http://jls.sagepub.com/content/28/4/441

Published by:
SAGE
http://www.sagepublications.com

Journal of Language and Social Psychology
28(4) 441–456

10.1177/0261927X09341950
http://jls.sagepub.com
SAGE

Reverse Linguistic Stereotyping: Measuring the Effect of Listener Expectations on Speech Evaluation

Okim Kang[1] and Donald L. Rubin[2]

Abstract

The linguistic stereotyping hypothesis holds that even brief samples of speech varieties associated with low-prestige groups can cue negative attributions regarding individual speakers. The converse phenomenon is *reverse linguistic stereotyping* (RLS). In RLS, attributions of a speaker's group membership trigger distorted evaluations of that person's speech. The present study established a procedure for ascertaining a proclivity to RLS for individual listeners. In addition to RLS, variables reflecting degree of multicultural involvement (e.g., proportion of friends who are nonnative speakers, amount of language study) predicted speech evaluations. Although the RLS measurement procedure outlined here requires more demanding administration than mere paper-and-pencil self-reports, it has the advantage of reflecting authentic RLS processes. Measuring individuals' RLS levels can help screen teachers, job interviewers, immigration officials, and others who are called on to make judgments about the oral proficiency of speakers of nonprestige language varieties.

Keywords

reverse linguistic stereotyping, speech evaluation

[1]Northern Arizona University, Flagstaff, AZ, USA
[2]University of Georgia, Athens, GA, USA

Corresponding Author:
Okim Kang, Northern Arizona University, Liberal Arts Bldg 18-Room 140, PO Box 6032, Flagstaff, AZ, 86011, USA
Email: okim.kang@nau.edu

The purpose of this study is to examine a particular kind of language-related stereotyping process that we call *reverse linguistic stereotyping* (RLS). Beginning with the groundbreaking work of Lambert and his colleagues (Lambert, Hodgson, Gardner, & Fillenbaum, 1960) and continuing now for nearly half a century (see review and prospect in Bradac, Cargile, & Hallett, 2001), the linguistic stereotyping hypothesis holds that even brief samples of speech varieties (e.g., dialects, genderlects, minority languages) associated with low-prestige groups can cue negative attributions regarding individual speakers. Thus, for example, even though attitudes toward Jamaican Creole are fairly positive among Jamaicans they regard an English speaker to be more intelligent than a Creole speaker on the average (Jamaican Language Unit, 2005). In other words, listeners attribute to individual members of a group the traits that they stereotypically ascribe to the group, and speech patterns are a major trigger to those attributional processes (Johnson, 2000; Lippi-Green, 1997).

RLS is the converse of the linguistic stereotyping hypothesis. In RLS, the speaker's language pattern is not the trigger to stereotyping processes but rather their object. In RLS, attributions of a speaker's group membership cue distorted perceptions of that speaker's language style or proficiency. Thus, Rubin and colleagues (see review in Rubin, 2002) have repeatedly documented that when listeners mistakenly *believe* they are listening to a nonnative speaker of English (NNS), they report *hearing* highly accented speech, and their listening comprehension significantly declines.

The sort of RLS effect to which this work points—that is, extending general judgments about social groups to evaluations of individual speakers' language proficiency—is corroborated by other research. Nguyen (1993), for example, concluded that inherent native speaker (NS) rater biases against certain nationalities renders valid standardized testing of oral proficiency unattainable for English language learners from those countries. More recently, Lindemann (2002, 2003) showed that such stereotypes materially affect listeners' communication behaviors. For example, U.S. undergraduate students reduced their question asking with instructors whom *they believed* to be of particular (negatively stereotyped) NNS backgrounds.

What accounts for the proclivity to engage in RLS? One early study found that among U.S. undergraduates, the tendency to engage in stereotyped listening of international instructors was inversely related to exposure to international instructors (Rubin & Smith, 1990). In other words, NS listeners who "stuck with" cross-cultural engagements were subsequently rewarded with less distorted listening outcomes. Thus, listeners' background characteristics that pertain to amount of contact with NNSs likely influence the perception process in evaluating NNS speech. And indeed, individuals unfamiliar with a particular variety of accented English generally perceive a higher degree of second language foreign accent than do those who are familiar with that particular variety (Jenkins, 2000; Thompson, 1991). In a similar vein, raters with previous experience teaching NNSs tend to rate NNS speech more positively than those with no such teaching experience (Barnwell, 1989).

Although a number of prior studies, then, identify the impact of apparent RLS on speech evaluations, none has attempted to directly measure the RLS construct. Moreover, there has been little study about a host of other issues related to the effect of listener expectation on speech evaluation. As a result, this study was guided by the following research questions:

Research Question 1: Can RLS be measured as a listener propensity?
Research Question 2: To what degree does measured propensity to linguistic stereotyping affect raters' judgments of NNS oral performances?
Research Question 3: To what degree do listeners' characteristics—especially indices of multicultural immersion—predict variance beyond that attributable to RLS in raters' judgments of NNS oral performances?

Reverse Linguistic Stereotyping and Listener Expectation

A number of matched-guise studies on language attitude have demonstrated that people make moral, intellectual, and aesthetic judgments of others based on language choice and accent alone. In matched-guise studies, listeners are typically asked to rate recorded speakers on a number of qualities, which may be divided into status-related qualities such as intelligence and ambition and solidarity-related qualities such as friendliness and likeability (e.g., Campbell-Kibler, 2007). The speaker variable (effects of voice quality) is controlled by having the same bilingual (or bidialectal) speaker record both language guises (Lambert et al., 1960). Many of the dimensions of evaluation measured in matched-guise studies touch on personality characteristics such as confidence and enthusiasm (Williams, 1976) and even quite extraneous traits such as physical height and attractiveness (Edwards, 1982; Seligman, Tucker, & Lambert, 1972).

Quite consequentially, these sorts of judgments may result in language-based discrimination. Students with "poor voices" are judged by teachers to be less intelligent than those with "good voices" (Seligman et al., 1972). Australians with "broad" accents are rated by potential employers as unsuitable for high-status jobs (Seggie, Smith, & Hodgins, 1986). Speakers of African American vernacular English are misinformed by landlords that there are no available apartments (Purnell, Idsardi, & Baugh, 1999). Similarly, NNSs may be denied raises or even fired by employers who claim they have poor language proficiency (Lippi-Green, 1997).

In the case of NNSs, issues of social attribution are compounded if the speakers are members of stigmatized groups and speak with stigmatized accents that index them as such. In the case of U.S. monolingual NSs' perceptions of Spanish-accented English, for example—a case in which NNSs are both out-group and low prestige—these NNSs were rated lower on measures of both status and solidarity (Ryan, Carranza, & Moffie, 1977; Ryan & Sebastian, 1980). In the same vein, NNSs are vulnerable to linguistic stereotyping when their second language proficiency is being evaluated. Proficiency judgments may have more to do with listeners' attitudes

about the speakers' ethnicity than with the speakers' actual intelligibility (Lippi-Green, 1997).

Rubin's (1992, 2002) work related to the RLS construct has shown that listener expectations based on speaker nationality can affect listener comprehension as well as social judgment. In a typical study in this paradigm, participants listened to 4 minutes of a tape-recorded lecture produced by a native speaker of standard American English. Some participants were lad to believe that they were listening to a North American NS instructor, whereas others were lead to believe that the instructor was an international NNS. Instructor ethnicity/nationality was operationalized by projecting a photograph of either a Caucasian model or an East-Asian model. The fabricated instructors were also assigned either an Anglo-Saxon name and home of origin in the United States, or a Chinese name and home of origin in China. Typically, U.S. NS undergraduates harbor certain negative expectations about East-Asian teaching assistants (Bresnahan & Kim, 1993; Fox & Gay, 1994) and therefore would be expected to derogate the speech of such instructors. And indeed, listeners who were exposed to the Chinese/NNS guise perceived more of a foreign accent and scored lower on a recall test than those who were exposed to the Caucasian/NS guise even though the audiotape they heard was exactly identical (standard American English). In other words, attributing social identity to the speaker affected listeners' processing and evaluation of the speech. The finding of RLS has been replicated a number of times (e.g., Rubin, Ainsworth, Cho, Turk, & Winn, 1999).

Effects of Listener Background Characteristics

Little is known about what individual differences predispose some listeners to RLS. Certain listener background variables—apart from any speaker performance factors—however, are known to predict evaluations of NNSs' performance. Diverse listener groups differ in judging learners' second language ability. In some studies, NNS judges were harsher than NS judges in evaluating NNS English (e.g., Fayer & Krasinski, 1987; Santos, 1988). When listeners from one language group share few phonological features with speakers from another language group, comprehension suffers (Deterding & Kirkpatrick, 2006). On the other hand, listeners from particular language backgrounds sometimes exhibit special tolerance for certain nonnative accents (Bent & Bradlow, 2003). For example, because of some phonological similarity among Chinese, Japanese, and Spanish, Chinese and Japanese listeners understand Spanish-accented English relatively well (Major, Fitzmaurice, Bunta, & Balasubramanian, 2002).

Training and experience as a language teacher may confer a lenient mind-set on listeners evaluating NNSs' speech (Barnwell, 1989; Galloway, 1980; Hadden; 1991). The degree of international contact and exposure to varieties of NNSs of English similarly may mitigate judgments of NNSs' speech. Derwing and Munro (1997) found that listeners' self-reported exposure to various accents predicted their success at language identification and correlated with their intelligibility (i.e., word

recognition) scores. It is commonly observed that interaction with speakers of specific World Englishes (e.g., Nigerian English or Bengali English) facilitates listeners' comprehension of those varieties (Clark & Garrett, 2004; Field, 2003; Gass & Varonis, 1984). That is, the more opportunity one has to listen to a particular accent, the easier it becomes to comprehend speakers of that specific accent. On the other hand, some studies (e.g., Powers, Schedl, Wilson-Leung, & Butler, 1999) found that no listener background variables were consistently related to raters' evaluations.

In summary, previous research indicates that judgments of accent and oral proficiency are susceptible to listener expectations based on the speaker's social identity. Even listening comprehension processes are vulnerable to social stereotypes. No previous research has directly measured the propensity for this RLS, however. Furthermore, the propensity to RLS is an individual difference that results from one's experience with speakers of different language varieties or from one's general multicultural exposure. Previous studies, though, have yielded mixed results with respect to the association of such listener background variables with leniency or stringency in judging NNSs and their speech. The present study, accordingly, describes a method for directly observing RLS and tests individual background variables that are likely determinants of that propensity.

Method

Participants

Usable data were collected from 158 individuals sampled deliberately for their diverse backgrounds. They were recruited by advertising in the campus and local newspaper and in world languages classes. They were recruited such that they could be expected to collectively vary across the rater background dimensions: (a) native English language speaker status (native/nonnative), (b) composite index of exposure to nonnative English-speaking friends and acquaintances, (c) formal training in language studies, and (d) experience in language teaching/tutoring. No participants with previous experience in standardized language-rating activities were selected. Participants were remunerated for their time.

Degree of participants' exposure to NNSs was indexed by listeners' self-reports of the number of hours spent with NNSs during a typical week. Linguistic sophistication was derived by summing (a) the number of college classes in linguistics, applied linguistics, or test of English as second language methods with (b) years of foreign language study. The amount of teaching/tutoring experience was determined by summing raters' teaching/tutoring experience in either English as a second language or foreign languages in weeks. The distribution of these rater background characteristics of interest is shown in Table 1. The total sample size of listeners yielded .80 statisical power for medium effect size, based on Gatsonis and Sampson's (1989) calcuation.

Table 1. Participant Characteristics on Multicultural Background Variables of Interest

Listener Status	*N*	Weeks Taught/ Tutored ESL/EFL (Avg.)	Linguistic/ TESOL Classes (Avg.)	Weekly Contact With NNSs (Avg.)
NS	102	78.86	3.12	29.67%
NNS	56	147.95	5.82	77.96%

Note: ESL = English as second language; EFL = English as first language; TESOL = test of English as second language; NS = native speaker; NNS = nonnative speaker.

Measuring Language Attitudes

To observe RLS in an American English context, one must measure listeners' language attitudes toward one putative speaker who is ascribed a Euro-American NS identity and also measure language attitudes toward another putative speaker who is ascribed an identity as a "foreign" NNS. Both putative "speakers," however, are just different guises for the same recorded voice. Research on language attitudes has used a variety of dependent variables to gauge the dimensions of perception whereby listeners judge speakers (Edwards, 1982). The Speech Evaluation Instrument (SEI) developed by Zahn and Hopper (1985) has been used in dozens of studies of language attitudes (e.g., Cargile, 2002; Dailey-O'Cain, 2000; Rubin, 1992; Rubin & Smith, 1990). It typically factor analyzes into three dimensions: (a) *superiority*, (b) *social attractiveness*, and (c) *dynamism*. Internal consistency reliability (Cronbach's alpha) was calculated for each of the three subscales separately for (a) the Euro-American guise and (b) for the East-Asian guise ratings. All six reliabilities were acceptable ($.93 \leq \alpha \geq .80$). Because the dynamism dimension does not necessarily index stigmatization—that is, members of stigmatized groups may score high on measures of confidence/enthusiasm (Williams, 1976)—dynamism ratings were excluded from further analysis in this study.

Interspersed among the SEI items in the present study were additional semantic differential items (e.g., see Kerlinger, 1973), 7-point bipolar scales. Each item posed polar opposite descriptions at either end of seven equal appearing intervals. Participants checked the ethnicity/nationality manipulations of the photographs ("Caucasian/European ethnicity, . . ., Oriental/Asian ethnicity") as well as the criterion measures. Altogether, participants rated the speakers and their speech on 35 semantic-differential items.

RLS Procedures

The data were collected in a series of face-to-face meetings with 8 to 15 participants each. Venues for each meeting were identically structured with the same audiotaped lecture and guised photographs. Each session was randomly assigned to present the

Asian guise first and the Euro-American guise second, or vice versa. After providing informed consent, participants completed the questionnaire items that yielded data about their background in multicultural exposure.

For the measure of RLS, participants heard different 4-minute sections of an audiorecording simulating a portion of a college lecture about galaxies, previously used in similar language and attitude research (e.g., Rubin & Smith, 1990). This lecture was selected because the general topic schema would likely be familiar to a wide range of listeners, whereas the specific information imparted would not be. Participants listened to the lecture segment once with a Caucasian face projected and once with an Asian face (counterbalanced for order). A distractor lecture segment was played between the two target listening tasks. For the distractor, all listeners heard a distinctly East-Asian speaker of intermediate intelligibility. Our brief informal interviews with participants, after the completion of the study, found that most raters either perceived the first voice as distinct from the third voice or reported being uncertain as to whether the voices were distinct. Participants rated the distractor speech sample as well as the two target speech samples.

The audiorecorded lecturer for both of the target listening passages was the same male speaker of standard American English, originally from a small town in Michigan, a teacher of speech communication who was acknowledged by peers to have a particularly clear speaking voice. However, the speaker was identified through fabricated photographs and dossiers as either an NNS Chinese international teaching assistant or an NS Euro-American teaching assistant. A manipulation check indicated that the East-Asian guise was perceived to be a "person of color" more so than was the Euro-American guise ($M_{\text{Asian}} = 5.95$, $M_{\text{Euro}} = 3.69$; $t(68) = 6.72$, $p < .000$).

Physical attractiveness can exert strong influence on social judgments (Riniolo, Johnson, Sherman, & Misso, 2006). To avoid confounding ethnicity with physical attractiveness, both male models were similarly dressed, were of similar size and hair style (i.e., dark-haired), and were photographed in the same setting and pose (standing in front of a whiteboard). Pretesting indicated no significant difference in perceived physical attractiveness between the two models. Despite these efforts to minimize average differences in physical attractiveness, individually perceived physical attractiveness was used as a covariate in subsequent analyses (see below).

The RLS procedure ultimately yielded two dependent variables that index distinct dimensions of linguistic stereotyping: *superiority RLS* and *social attractiveness RLS.* Superiority and social attractiveness, in turn, are derived from the SEI. The two RLS scores were indexed by subtracting speech evaluations accorded to the East-Asian guise from those accorded to the Euro-American guise for each of the two dimensions separately. To subtract out the effects of listener judgments about the speakers' physical attractiveness, the actual values used in these calculations were the unstandardized residuals from regression of two SEI scales on the values of the physical attractiveness item.

Criterion Measures

Listening comprehension. A cloze test was adopted from previous research in this domain (e.g., Rubin & Smith, 1990) to measure listening comprehension. A cloze test gives listeners a short text with blanks and asks them to fill in the blanks. In this study, participants were presented with a written transcript of the lecture they had heard on audiotape. Approximately every seventh word was deleted save for the first sentences, which were kept intact. There were 52 blanks out of 410 words of the text script. Only exact recall was scored as correct.

Teaching quality ratings. The teaching quality rating scale was composed of six semantic differential items. The scale was an extension of the four items used in earlier studies of undergraduates' responses to international teaching assistants (e.g., Rubin et al., 1999). Examples of the questions were "effective teacher or ineffective teacher" and "qualified or unqualified." The internal consistency reliability coefficient of this scale was marginally acceptable with .70. Accordingly, ratings on the nine instructional competence items were summed into a single scale measure.

Accent standardness ratings. The accent standardness rating scale was a single item measure (i.e., foreign accent or American accent). A similar version of this rating was used in Derwing and Munro (1997) and Rubin (1992).

Data Analysis

Data were analyzed through separate stepwise linear regressions for each of the three criterion (dependent) variables: (a) listening comprehension, (b) teaching quality ratings, and (c) accent standardness ratings for Asian guise and Caucasian guise speakers. The two RLS measures—(a) superiority RLS and (2) social attractiveness RLS—were entered as predictors in the first step of the regression, because ascertaining a proclivity to RLS for individual listeners was the primary purpose of this study. Subsequently, four additional background characteristic variables were entered as predictors in the second step. These listener background predictors included (c) dummy-coded native/nonnative English language speaker status, (d) composite index of exposure to nonnative English-speaking friends and acquaintances, (e) linguistic sophistication (formal training in language studies and linguistics), and (f) experience in language teaching and tutoring.

Results

Table 2 shows the zero-order correlations among the two RLS measures and four background predictor variables. As Table 2 indicates, collinear relations among them are not very strong. As might be expected, the strongest bivariate correlation among these seven variables is between the NNS status variable and exposure to NNS variable ($r = .49$).

Table 2. Correlations Among Two Reverse Linguistic Stereotype and Four Listener Background Variables

	Social Attractiveness RLS	NNS status	Linguistic Sophistication	Teaching Experience	Amount of Contact With NNS
Superiority RLS	.18*	.02	–.01	–.13	.11
Social attractiveness RLS		.13	–.00	–.03	.04
NNS status			.11	.12	.49**
Linguistic sophistication				.35**	–.02
Teaching experience					.01

Note: RLS = reverse linguistic stereotyping; NNS = nonnative speaker.
*$p < .05$, two-tailed. **$p < .01$, two-tailed.

Table 3. Multiple Regression of RLS Factors and Listener Characteristics on Listener Comprehension

	Standardized Coefficients (β)	*t* Value	*p* Value	Part Correlation
Superiority RLS	–.12	–1.51	.134	–.12
Social attractiveness RLS	–.18	–2.10	.037	–.18
NNS status	–.30	–3.35	.000	–.30
Linguistic sophistication	.17	2.03	.044	.17
Teaching experience	.13	1.50	.136	.13
Amount of contact with NNS	.15	1.65	.102	–.08

Note: RLS = reverse linguistic stereotyping; NNS = nonnative speaker. Step 1. $R^2 = .11$, $F(2, 151) = 2.86$, $p < .05$, adjusted $R^2 = .08$; Step 2. $R^2 = .23$, $F(7, 145) = 3.97$, $p < .01$, adjusted $R^2 = .18$.

Of considerable interest, no regression model attained statistical significance when listeners were responding to the NS Euro-American/Caucasian guise. Of course, though one must be cautious in inferring conclusions from a lack of statistical significance, the meaningfulness of the RLS construct would have been undermined had any of the predictors tested here affected judgments of a standard American English speaker.

In contrast, regression analyses for listeners' responses to the NNS East-Asian guise revealed statistically significant impact of several predictors.

Comprehension Scores

The multiple regression for comprehension scores is summarized in Table 3. Approximately 11% of the variance in the comprehension scores was explained by the 2 RLS predictor variables, and an additional 12% was contributed by the four listener background predictors.

Table 4. Multiple Regression of RLS Factors and Listener Characteristics on Teaching Quality Ratings

	Standardized Coefficients (β)	*t* Value	*p* Value	Part Correlation
Superiority RLS	–.27	–2.36	.020	–.21
Social attractiveness RLS	–.22	–2.12	.036	–.18
NNS status	–.01	–.14	.892	–.01
Linguistic sophistication	.09	.716	.475	.06
Teaching experience	.18	2.12	.036	.18
Amount of contact with NNS	.17	2.01	.051	.16

Note: NNS = RLS = reverse linguistic stereotyping; nonnative speaker. Step 1. $R^2 = .12$, $F(2, 151) = 2.80$, $p < .05$, adjusted $R^2 = .07$; Step 2. $R^2 = .22$, $F(7, 132) = 2.61$, $p < .05$, adjusted $R^2 = .14$.

One of the two RLS measures, *social attractiveness*, contributed inversely to listening comprehension. In other words, the propensity to RLS on this dimension of language attitude indeed resulted in lower comprehension when listening to the East-Asian guise. Additionally, NNS status revealed a negative regression coefficient. NSs (coded as 0) exhibited higher listening comprehension for the Asian guise speaker than did NNSs (coded as 1). In addition, linguistic sophistication was positively associated with comprehension of Asian guise's speech. None of the other listener trait variables exerted statistically significant effects on this dimension of NNS speech evaluation.

Teaching Quality Ratings

The multiple regression of teaching quality ratings is summarized in Table 4. Approximately 11% of the variance in this criterion variable was explained by the two RLS predictor variables, with an additional 10% contributed by the four listener background variables.

Both RLS measures, *superiority RLS* as well as *social attractiveness RLS*, inversely predicted rated NNS teaching quality. Propensity to RLS on these two dimensions of language stereotyping was associated with negative ratings of NNS teaching performance. English as second language/English as first language (ESL/EFL) teaching experience was directly proportional to listeners' rating of the Asian guise's teaching quality. None of the other listener trait variables exerted statistically significant effects on this dimension of NNS speech evaluation.

Accent Standardness Ratings

The multiple regression of accent standardness ratings is summarized in Table 5. Approximately 8% of the variance in this outcome variable was explained by the two

Table 5. Multiple Regression of RLS Factors and Listener Characteristics on Accent Standardness Ratings

	Standardized Coefficients (β)	*t* Value	*p* Value	Part Correlation
Superiority RLS	–.39	–2.62	.011	–.28
Social attractiveness RLS	–.26	–1.64	.105	–.17
NNS status	–.04	–.34	.699	–.04
Linguistic sophistication	.01	.054	.957	.01
Teaching experience	.05	.35	.728	.04
Amount of contact with NNS	.21	1.74	.085	.19

Note: RLS = reverse linguistic stereotyping; NNS = nonnative speaker. Step 1. $R^2 = .09$, $F(2, 151) = 3.88$, $p < .05$, adjusted $R^2 = .06$; Step 2. $R^2 = .18$, $F(7, 132) = 2.61$, $p < .05$, adjusted $R^2 = .13$.

RLS predictor variables, with an additional 9% contributed by the four listener background variables.

Superiority RLS inversely predicted perceived standardness of NNSs' accent. That is, propensity to RLS on this dimension resulted in listeners reporting that the NNS guise spoke with a nonstandard accent. (Recall that listeners were in every case listening to a standard American English speaker.) None of the other listener trait variables exerted statistically significant effects on perceived accent standardness.

Discussion

Nonnative speakers of English are often subjected to evaluations of their spoken English that have profound consequences for their education, employment, and even citizenship. However, NS judgments of NNS speech are notoriously biased. NS listeners often hear what they expect to hear rather than accurately perceive NNS speech. And what they expect to hear is often quite unsatisfactory. RLS is that very process of evaluating a person's speaking performance based on stereotypes associated with the speaker's social identity. The primary purpose of this study was to ascertain the proclivity to engage in RLS for individual listeners and to investigate the proportion of variance in ratings of NNS speech attributable to listener RLS and to listener background characteristics related to multicultural exposure.

Listener RLS as well as listener background factors did contribute substantial variance to ratings of NNSs' oral performance. In rating a speaker with attributed NNS/East-Asian identity, 18% to 23% of the variance in all three listening outcomes—listening comprehension, rated instructional quality, and perceived accent standardness—were attributable collectively to listener RLS and background factors. Approximately 9% to 12% of the variance in NNSs' oral performance ratings was specifically attributable to the RLS dimensions: superiority RLS and social attractiveness RLS. These findings confirm previous conclusions that ratings of

speaking skills are susceptible to rater expectation and stereotype (Bradac et al., 2001; Piché, Michlin, Rubin, & Sullivan, 1977; Rubin, 2002).

The first research question of this study queried how RLS can be measured as a listener propensity. Perhaps the most important contribution of this article is to describe a procedure by which individuals' propensity to engage in RLS can be directly observed. In this procedure, listeners are exposed to speech samples derived from the identical speaker of the standard dialect. In one listening guise, a prestige NS identity is attributed to this speaker, whereas in another guise, a stigmatized "foreign" NNS identity is attributed. The degree to which those social attributions differentially affect ratings and comprehension of the speech indexes propensity to RLS.

With regard to the second research question (i.e., degree to which measured propensity to RLS affects listeners' judgments of NNS oral performances), both RLS dimensions proved potent. This study's findings are compatible with the view that perceptions of speaker accent are distorted by listeners' propensity to engage in RLS. Raters who tended to judge the speech of NNSs in an especially stereotypical manner rated the East-Asian guise as particularly accented (though it was actually spoken by an expert NS).

Listeners marked by superiority RLS "heard" more accented speech when they were mistakenly made to believe they were listening to a native Chinese instructor. Listeners with marked social attractiveness RLS (considering NNSs as unfriendly, cold, hostile, dishonest) actually suffered comprehension loss when they thought they were listening to a native Chinese instructor. And listeners marked by both dimensions of RLS had low evaluations of that instructor's teaching prowess. Moreover, the fact that none of the regression models in the Caucasian guise emerged as statistically signficant was striking, because the Asian guise and the Caucasian guise incorporated audiorecordings of the very same speaker. As might be supposed, RLS is a process that affects our perceptions of NNS speakers, but not of NS speakers.

The last research question queried about the contribution of listeners' characteristics to their comprehension and judgments of NNS oral performances. One would expect NSs of English to have better listening comprehension of English than NNSs, and indeed, NS in this study had higher listening comprehension (cloze test scores) for the Asian guise speaker than did NNSs. (Surprisingly, though, cloze test scores for the two groups of listeners did not differ in the case of the Euro-American guise.) Formal course work in linguistics and the number of foreign language courses taken showed positive impact on comprehension of Asian guise's speech. This finding indicates that the higher language sophistication was indexed the better was raters' comprehension of NNS's speech. If a high degree of education in linguistics and language equates with high comprehension of NNS speech, this finding suggests that U.S. listeners would benefit by becoming more actively involved in learning additional languages.

As far as the teaching background was concerned, ESL/EFL teaching experience was directly proportional to listeners' rating of the Asian guise's teaching quality.

Listeners who have taught or tutored English in the past seemed to be more compassionate raters of NNSs' oral performances. This result is consistent with Barnwell's (1989) study, which reported that nonteaching raters were relatively harsher than a group of teachers.

The last background variable, exposure to NNSs, was not statistically associated with any of the three listening outcomes (though it might be mentioned in passing that $p = .051$ for teacher quality ratings). As has been found repeatedly in research on the contact hypothesis, mere contact between groups does not guarantee reduction of stereotypes and prejudice (see review in Rubin & Lannutti, 2001). Rather, intergroup contact must be facilitated in very specific ways if it is to have that salubrious result.

Certainly limitations must be acknowledged in weighing the conclusions of this article. We believe that high construct validity inheres in this procedure for observing propensity to engage in RLS, but resulting scores have not been examined for test/retest reliability. We do not know if the propensities it reveals are stable over time and context. Indeed, context of the rating task may have strong ramifications. Participants in this study were recruited specifically to engage in a rating activity and therefore their mindsets may have been directed toward issues of fairness and discernment. No doubt the degree of RLS observed in this study underestimates RLS processes that take place in less self-conscious evaluative contexts, such as when U.S. undergraduates criticize the language and teaching competence of their current international teaching assistants.

Language is a living object buffeted by human emotions and perceptions. RLS is an ongoing act of social discrimination in which individuals' language use is misjudged and misunderstood by virtue of listeners' stereotypes of speakers' social identities. The findings of this study imply that listeners who tend to engage in RLS also tend to find NNSs' speech more difficult to understand, more heavily accented, and they also tend to derogate such speakers' teaching performance. Results of this study comport with consistent findings (e.g., Rubin, 2002) of about 12% decrement in listening comprehension when U.S. undergraduates are led to believe they are listening to an international teaching assistant even though they are in fact listening to an NS.

Because language judgments have tangible impact on individuals' opportunities for education, for career advancement, and even for civil rights, RLS is of more than just scholarly interest. In a practical sense, perhaps a procedure such as that described here can be used to help screen potential raters of high stakes tests, teachers and job interviewers, and even immigration officials, so that judgments about NNSs' performances can become more accurately based on trait-relevant factors such as true speech comprehensibility.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interests with respect to their authorship or the publication of this article.

Funding

This research was partially supported by a dissertation grant in aid from the Spaan Foundation at University of Michigan and also by a research contract with the TOEFL® Committee of Examiners, Educational Testing Service. Conclusions are solely those of the authors.

References

Barnwell, D. (1989). Naïve native speakers and judgments of oral proficiency in Spanish. *Language Testing, 6*, 152-163.

Bent, T., & Bradlow, A. (2003). The interlanguage speech intelligibility benefit. *Journal of the Acoustical Society of America, 114*, 1600-1610.

Bradac, J. J., Cargile, A. C., & Hallett, J. S. (2001). Language attitudes: Retrospect, conspect, and prospect. In W. P. Robinson & H. Giles (Eds.), *The new handbook of language and social psychology* (pp. 137-158). Chichester, UK: John Wiley.

Bresnahan, M. I., & Kim, M. S. (1993). Factors of receptivity and resistance toward international teaching assistants. *Journal of Asian Pacific Communication, 4*, 1-12.

Campbell-Kibler, K. (2007). Accent, (ING), and the social logic of listener perceptions. *American Speech, 82*, 32-64.

Cargile, A. C. (2002). Speaker evaluation measures of language attitudes: Evidence of information-processing effects. *Language Awareness, 11*, 178-191.

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign accented English. *Journal of the Acoustical Society of America, 116*, 3647-3658.

Dailey-O'Cain, J. (2000). The sociolinguistic distribution of and attitudes toward focuser like and quotative like. *Journal of Sociolinguistics, 4*, 60-80.

Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition, 19*, 1-16.

Deterding, D., & Kirkpatrick, A. (2006). Intelligibility and an emerging ASEAN English lingua franca. *World Englishes, 25*, 391-410.

Edwards, J. R. (1982). Language attitudes and their implications among English speakers. In E. B. Ryan & H. Giles (Eds.), *Attitudes towards language variation* (pp. 20-33). London: Edward Arnold.

Fayer, J. M., & Krasinski, E. (1987). Native and nonnative judgments of intelligibility and irritation. *Language Learning, 37*, 313-326.

Field, J. (2003). The fuzzy notion of "intelligibility": A headache for pronunciation teachers and oral testers. *IATEFL Special Interest Groups Newsletter,* 34-38.

Fox, W. S., & Gay, G. (1994). Functions and effects of international teaching assistants. *Review of Higher Education, 18*, 1-24.

Galloway, V. B. (1980). Perceptions of the communicative efforts of American students of Spanish. *Modern Language Journal, 64*, 428-433.

Gass, S. M., & Varonis, E. M. (1984). The effect of familiarity on the comprehensibility of nonnative speech. *Language Learning, 34*, 65-89.

Gatsonis, C., & Sampson, A. R. (1989). Multiple correlation: Exact power and sample size calculations. *Psychological Bulletin, 106*, 516-524.

Hadden, B. (1991). Teacher and nonteacher perceptions of second-language communication. *Language Learning, 41*, 1-24.

Jamaican Language Unit. (2005). *The language attitude survey of Jamaica*. Mona, Jamaica: University of the West Indies Department of Language, Linguistics & Philosophy. Retrieved September 2, 2008, from http://www.mona.uwi.edu/dllp/jlu/projects/report%20for%20language%20attitude%20survey%20of%20jamaica.pdf

Jenkins, S. (2000). Cultural and linguistic miscues: A case study of international teaching assistant and academic faculty miscommunication. *International Journal of Intercultural Relations, 24*, 477-501.

Johnson, F. L. (2000). Speaking culturally: Language diversity in the United States. Thousand Oaks, CA: Sage.

Kerlinger, F. N. (1973). *Foundations of behavioral research*. New York: Holt, Rinehart & Winston.

Lambert, W. E., Hodgson, R. C., Gardner, R. C., & Fillenbaum, S. (1960). Evaluational reactions to spoken language. *Journal of Abnormal and Social Psychology, 60*, 44-51.

Lindemann, S. (2002). Listening with an attitude: A model of native-speaker comprehension of non-native speakers in the United States. *Language in Society, 31*, 419-441.

Lindemann, S. (2003). Koreans, Chinese, or Indians? Attitudes and ideologies about non-native English speakers in the United States. *Journal of Sociolinguistics, 7*, 348-364.

Lippi-Green, R. (1997). *English with an Accent: Language, ideology, and discrimination in the United States*. New York: Routledge.

Major, R., Fitzmaurice, S., Bunta, F., & Balasubramanian, C. (2002). The effects of nonnative accents on listening comprehension: Implications for ESL assessment. *TESOL Quarterly, 36*, 173-190.

Nguyen, B. B. D. (1993). Accent discrimination and the test of spoken English: A call for an objective assessment of the comprehensibility of nonnative speaker. *Asian Law Journal, 81*, 1324-1361.

Piché, G. L., Michlin, M. L., Rubin, D. L., & Sullivan, A. (1977). Effects of dialect-ethnicity, social class, and quality of written compositions on teachers' subjective evaluations of children. *Communication Monographs, 44*, 60-72.

Powers, D. E., Schedl, M. A., Wilson-Leung, S. W., & Butler, F. A. (1999). Validating the revised Test of Spoken English against a criterion of communicative success. *Language Testing, 16*, 339-425.

Purnell, T., Idsardi W., & Baugh, J. (1999). Perceptual and phonetic experiments on American English dialect identification. *Journal of Language and Social Psychology, 18*, 10-30.

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

Rubin, D. L. (1992). Nonlanguage factors affecting undergraduate's judgments of nonnative English-speaking teaching assistants. *Research in Higher Education, 33*, 511-531.

Rubin, D. L. (2002). Help! My professor (or doctor or boss) doesn't talk English! In J. Martin, T. Nakayama, & L. Flores (Eds.), *Readings in intercultural communication: Experiences and contexts* (pp. 127-137). Boston: McGraw-Hill.

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

Rubin, D. L., & Lannutti, P. (2001). Frameworks for assessing contact as a tool for reducing prejudice. In V. H. Milhouse, M. K. Asante, & P. O. Nwosu (Eds.), *Transcultural realities: Interdisciplinary perspectives on cross-cultural relations* (pp. 313-326). Thousand Oaks, CA: Sage.

Rubin, D. L., & Smith, K. (l990). Effects of accent, ethnicity, and lecture topic on undergraduates' perceptions of non-native English speaking teaching assistants. *International Journal of Intercultural Relations*, *14*, 337-353.

Ryan, E. B., Carranza, M. A., & Moffie, R. W. (1977). Reactions toward varying degrees of accentedness in speech of Spanish-English. *Language and Speech, 20*, 267-273.

Ryan, E. B., & Sebastian, R. J. (1980). The effects of speech style and social class background on social judgments of speakers. *British Journal of Social and Clinical Psychology, 19*, 229-233.

Santos, T. (1988). Professors' reactions to the writing of nonnative-speaking students. *TESOL Quarterly, 22*, 69-90.

Seggie, I., Smith, N., & Hodgins, P. (1986). Evaluations of employment suitability based on accent alone: An Australian case study. *Language Sciences, 8*, 129-140.

Seligman, C. R., Tucker, G. R., & Lambert, W. E. (1972). The effects of speech style and other attributes on teachers' attitudes toward pupils. *Language in Society, 1*, 131-142.

Thompson, I. (1991). Foreign accents revisited: the English pronunciation of Russian immigrants. *Language Learning, 41*, 177-204.

Williams, F. (1976). *Explorations in the language attitudes of teachers*. Rowley, MA: Newbury House.

Zahn, C. J., & Hopper, R. (1985). Measuring language attitudes: The speech evaluation instrument. *Journal of Language and Social Psychology, 4*, 113-124.

Bios

Okim Kang (PhD, University of Georgia, 2008) is an assistant professor of applied linguistics at Northern Arizona University. Her research focuses on second language pronunciation, oral language proficiency assessment, speech perception and language attitudes, and world Englishes.

Don L. Rubin (PhD, University of Minnesota, 1978) is professor emeritus in the Departments of Speech Communication, Language and Literacy Education, and Linguistics at the University of Georgia. He is a senior research associate at the institution's Center for Health and Risk Communication. His research focuses on intercultural communication, oral language assessment, language and social identity, and health literacy.