

Sara Kennedy* and Pavel Trofimovich

First- and final-semester non-native students in an English-medium university: Judgments of their speech by university peers

Abstract: By the end of their studies, non-native speakers of English studying at English-medium universities have had several years of exposure to English in that setting. Do non-native students, particularly those enrolled in non-language-related programs, show different levels of second language (L2) speaking ability in their final semester of studies than non-native students in their first semester, as judged by other students in the university community? In this exploratory cross-sectional study, two matched groups of L2 English university students in their first or final semester of study in non-language-related programs ($N = 20$) were recorded in mock job interviews. The students were rated by two groups of raters for accentedness, comprehensibility, fluency, and communicative effectiveness. Both rater groups were university students; one group was from diverse academic programs, while the other group was studying human resource management (HRM). Although the first- and final-semester L2 English students differed in how long they had studied in English, no significant difference in ratings between first- and final-semester students was found. However, the two rater groups differed in how they rated accentedness and comprehensibility, suggesting that the nature of listeners' previous academic experience (e.g., with HRM) influences their judgments. The use of holistic rating scales to evaluate L2 speech is discussed, as well as the relationship between the nature of language exposure and the performance of the student and rater groups.

Keywords: university students, English as a second language (ESL), speaking, speech ratings, language exposure

***Corresponding author: Sara Kennedy:** Concordia University, Department of Education, LB-579, 1455, boul. de Maisonneuve O., Montréal, QC, H3G 1M8, Canada.
E-mail: sara.kennedy@education.concordia.ca

Pavel Trofimovich: Concordia University, Department of Education, LB-579, 1455, boul. de Maisonneuve O., Montréal, QC, H3G 1M8, Canada. E-mail: pavel.trofimovich@concordia.ca

1 Introduction

For numerous non-native students in North American post-secondary institutions (henceforth, NN students), their time at university or college is challenging not simply because of the content of their courses, but because they are studying in a second or additional language (L2). Many of these students have lived only a short period of time in the L2 environment and have never taken subject-matter courses (e.g., geography) in the L2 (Cheng et al. 2004). To succeed in a post-secondary setting, NN students need to master not only the course content but also the necessary L2 skills. Subject-matter instruction in an L2 can be an effective means of improving students' L2 skills (particularly their L2 speaking ability), provided that the instruction includes a strong focus on L2 learning (Lyster 2007). Although NN students indeed learn subject matter through the L2, most university or college degree programs (save those devoted to language for specific/academic purposes) do not target L2 skills; students are expected to have already developed those skills when they graduate.

Typically, the L2 skills of NN students are formally assessed before and sometimes upon admission. However, NN students' L2 skills, especially in speaking, are most *frequently* assessed by members of the university or college community, who repeatedly encounter these students. Almost none of these community members have been trained to assess L2 speaking ability; nonetheless, their impressions of NN students' speaking can have significant consequences for students, both academically and socially (Kang and Rubin 2009). For example, instructors or classmates may avoid interacting with an NN student whom they consider difficult to understand. Importantly, naïve (untrained) listeners' judgments of L2 speech may be substantially affected by listeners' previous experience or education (Bradlow and Bent 2008; Kennedy and Trofimovich 2008). For instance, some university students study human resource management, including employee performance in the workplace. The nature of these students' studies may influence the ways in which they assess the L2 speaking ability of NN students. The goal of this exploratory study was therefore twofold: first, to investigate to what extent members of a university community perceive differences in L2 speaking ability between NN students at various stages of academic studies and, second, to examine whether students who are enrolled in a human resource management program rate NN students' L2 speaking ability differently than other listeners do.

1.1 L2 experience and the development of L2 speech in adults

Much research has been conducted on the relationship between adults' development of L2 speech and their L2 experience. Researchers have measured L2 speech both acoustically and through scalar judgments, investigating aspects such as vowel quality, voice onset time, strength of foreign accent, and fluency. L2 experience has typically been operationalized as the length of time L2 speakers have spent as adults living in an L2 setting, usually referred to as length of residence (LOR). Results from such research have been mixed. In some studies, adult L2 speakers with greater LOR were more nativelike in their speech than adult L2 speakers with shorter LOR (e.g., Flege and Fletcher 1992; Flege et al. 1999). In other research, adult speakers did not show differences in L2 speech as a function of LOR (e.g., Flege et al. 2006).

1.2 Adults' L2 experience and development of L2 speech in academic contexts

Research on the link between the L2 speech and L2 experience of adult NN students in academic contexts has almost entirely targeted learners in study-abroad programs. In such programs, these learners typically focus on the L2 itself or study in a language-related program (e.g., Freed 1995). Early research showed clear gains over time for study-abroad students in the development of L2 speech in various languages (Freed 1998). However, in more recent investigations, outcomes for L2 speech development have been mixed. For example, study-abroad students living and studying in Spain for one semester significantly improved in oral fluency measures and oral proficiency ratings (Segalowitz and Freed 2004) but made no gains in the pronunciation of particular Spanish consonants (Díaz-Campos 2004). Similarly, study-abroad students living in France for one term did not make any significant gains over time in oral fluency (Freed et al. 2004).

While they are relevant, findings from study-abroad research are not fully applicable to the learning conditions or outcomes of NN students in North America, whose courses and instructors as a rule do not focus on L2 skills. Of the existing research on NN students, most studies have targeted either students' growing acculturation and socialization into the L2 academic discourse environment (e.g., Cheng and Fox 2008; Zappa-Hollman 2007) or students' own perceptions of performing academic tasks and maintaining social contact in the L2 (e.g., Ferris 1998). One consistent finding has been NN students' self-reported difficulty in

communicating with their peers, other university students. As one student said, “Because English is not my language, I have to form entire sentences before uttering them . . . My Canadian colleagues often call me ‘thoughtful’ . . . It’s an expression of how difficult [it is] to converse with me” (Lee 2006: 82).

Only two studies from the University of Ottawa in Canada have included specific measures of NN students’ speech development over time (Burger and Chrétien 2001; Ready and Wesche 1992). The University of Ottawa is bilingual (English-French), and many undergraduate subject-matter courses are offered in both languages. Students who are not native speakers of the language of instruction can also take weekly adjunct L2 courses focusing particularly on reading, writing, listening, and speaking skills using subject-matter content. Ready and Wesche (1992) found that 90% of the NN French and English students in various adjunct L2 courses reported that their speaking ability had improved after two semesters. Burger and Chrétien (2001) showed that over an eight-month period both the NN French and English students taking an introductory psychology course along with an adjunct L2 class significantly improved in performance in at least one of the two speaking tasks (elicited imitation, discussion).

Apart from research in the study abroad domain, the two studies above seem to be the only ones to have explored NN university students’ L2 speech development in *second language* academic contexts. This differs from contexts for Content and Language Integrated Learning (CLIL), where the language of instruction is not commonly used in the surrounding environment. In several studies, the development of the language of university students in CLIL contexts has been measured longitudinally (e.g., Rauto 2006). Nonetheless, even in CLIL contexts, differences in students’ speech over time or compared to non-CLIL students has been studied only for primary or secondary school (e.g. Lasagabaster 2008). These studies generally show that when students receive language instruction which is tied to a subject-matter course and which includes listening and speaking activities, students’ speaking ability can improve. However, in second language contexts, many NN students either do not receive this kind of instruction at university or receive instruction which focuses only on reading and writing; NN students consistently report difficulties communicating with fellow students. It is thus important to explore whether the L2 speaking skills of NN students develop simply by virtue of studying at university. Thus, the first objective of the current cross-sectional study, designed as a preliminary step towards a longitudinal investigation, was to compare the L2 speech of first- and final-semester NN students in order to identify possible differences in L2 speaking ability for students at the beginning and the end of their university studies.

1.3 Using ratings to measure L2 speech

As mentioned above, L2 speech has typically been investigated using two different but complementary approaches: objective measurement of acoustic and temporal properties of speech and human rating. The reliability and accuracy of human rating has sometimes been questioned (Kang and Rubin 2009), particularly with respect to high-stakes assessment of L2 speech (Carey et al. 2011). However, as mentioned above, NN students (and other NN speakers in L2 settings) are most frequently judged *impressionistically*, by listeners who are not trained in assessing L2 speech. Unlike trained raters, who are trained to use specific criteria or standards for particular assessments, untrained raters rely on their own implicit or explicit criteria and standards in evaluating L2 speech. Therefore, compared to trained raters, untrained raters are more representative of how members of the general community might react to L2 speakers. Untrained raters also appear to be highly consistent in their reactions to L2 speech, with rater consistency values (Cronbach alpha) typically ranging between .94 and .99 for scalar ratings of speech using minimal descriptors (e.g., Derwing et al. 2004; Kang et al. 2010). Clearly, untrained raters tend to rate in very similar ways to each other.

Untrained raters may use rating scales in similar ways, but can they use them to differentiate between different levels of L2 speaking ability? Southwood and Flege (1999), who compared untrained English-speaking raters' use of several scaling techniques (including a 7-point Likert-type scale) for rating L2 accentedness, found that the raters were able to differentiate between at least seven levels of accentedness. The ability to differentiate between degrees of speaking ability was also evident in Derwing et al. (2008), where untrained native-speaking raters showed sensitivity to improvement in comprehensibility (ease of listeners' understanding) and fluency (smoothness of speech) through scalar ratings of the English speech samples of Slavic and Mandarin speakers, collected at three points over 20 months. In other research, untrained undergraduate students using Likert-type scales behaved similarly to trained raters in evaluating L2 speakers at various levels of proficiency (Bridgeman et al. 2012). Taken together, these findings show that untrained raters using Likert-type scales can perceive differences in L2 speakers' ability both over time and across speaker groups.

Although ratings appear to be a reliable source of information about L2 speech, to the best of our knowledge, there is no research focusing on the effects of listener background and experience on L2 speech ratings in an academic context. Members of university or college communities (e.g., teachers, students, support staff) interact with NN students on a regular basis, such that their judgments of L2 speech may have significant implications for the frequency and nature of their interactions with NN students (see Kang and Rubin 2009). Similar to

untrained raters in research settings, most members of academic communities do not have specialized training in assessing L2 speech. Therefore, they draw on their own criteria and standards when forming impressions of NN students. However, some members of academic communities with particular academic experiences (e.g., students in human resource management programs who are learning ways to evaluate employee performance) may develop similar criteria in judging L2 speech. This raises the question of whether students in human resource management programs judge L2 speech differently from other student raters. Therefore, the second objective of this exploratory study was to compare student raters in human resource management programs and student raters enrolled in other academic programs to determine if the two groups evaluate NN students' L2 speech differently. The overall objective here was to explore whether raters with specialized academic experiences (i.e., studies in human resource management), as potential future employers of NN students, would differ from other raters in their reactions to NN students' L2 speech.

1.4 The current study

As discussed in the introduction, the issue of possible differences in the L2 speech of NN students at the beginning of their post-secondary studies and NN students at the end of their studies has not yet been explored. These differences include not only those which are acoustically observable but also those which are noticeable to other university or college students with no specialized training in evaluating language proficiency. Moreover, it is unclear whether students in human resource management might evaluate the L2 speech of NN students differently than other students do. Therefore, in the current exploratory cross-sectional study, the speech of two groups of NN students was rated: NN students in their first semester and NN students in their final semester of studies. The two rater groups were students in human resource management programs and students from other academic disciplines.

NN students' speech was evaluated along four dimensions: accentedness, comprehensibility, fluency, and communicative effectiveness. Accentedness is here defined as a listener's judgment of how closely a speaker's speech approximates a native speaker norm. As a construct, accentedness is different from comprehensibility, a listener's perception of how easy or difficult it is to understand a speaker (Derwing et al. 1998). Fluency refers to a listener's perception of how fluidly and effortlessly speech is produced, that is, spoken without undue pauses, hesitations, or dysfluencies such as false starts and repetitions (Derwing et al., 2004). Because an L2 speaker may be considered "good" without necessarily be-

ing rated highly on accentedness, comprehensibility, or fluency, the inclusion of communicative effectiveness as a rating criterion allows raters to give their global impression of a speaker's ability to communicate in an L2. In the context of this study, communicative effectiveness refers to a listener's perception of speakers' ability "to get [their] message across, . . . to get people's attention, . . . to communicate" (Lehtonen and Karjalainen 2008: 498). The following two research questions were addressed:

1. Do university students with no training in assessing L2 speech perceive differences in the L2 speech of NN students at the end of their academic studies compared to NN students at the beginning of their studies?
2. Do university students studying human resource management evaluate NN students differently than do students from other academic programs?

2 Material and methods

2.1 Participants

2.1.1 Speakers

This study was cross-sectional in design, involving two groups of 10 non-native English speakers who were full-time students at an English-medium university in Quebec, Canada ($N = 20$). They were part of a larger group of students participating in research on NN students in university settings. The students had responded to notices inviting them to practice interviewing for jobs as part of a research study. The groups in the current study were selected from the larger pool of participants, following the matched subgroup method used in cross-sectional research (Flege et al. 1999). More specifically, the two participant groups were matched for a number of participant background variables (i.e., age, gender, undergraduate/graduate status, and age of arrival in Canada), which minimized possible extraneous effects of these variables on study outcomes. Table 1 shows students' relevant background information. The 20 students came from a variety of first language (L1) backgrounds, the most common being Mandarin ($n = 5$) and Arabic ($n = 4$). They were registered in different academic programs (education, economics, computer science, film studies), but most students in either group (14 out of 20, or 70%) were engineering or commerce students. Thus, the two groups were matched for several background characteristics but clearly differed in terms of timing of their studies (first semester vs. final semester), a variable of principal interest here.

Table 1: NN student characteristics

Measure	First semester (<i>n</i> = 10)			Final semester (<i>n</i> = 10)		
	<i>M</i>	<i>SD</i>	<i>Range</i>	<i>M</i>	<i>SD</i>	<i>Range</i>
Age (years)	24.7	3.9	19–33	27.2	4.7	21–33
Age of arrival (years)	22.7	5.0	18–31	23.0	6.3	18–30
Length of study (months)	1.2	0.7	1–2	28.3	4.1	24–32
Gender	5 female, 5 male			6 female, 4 male		
L1	Mandarin (3), Spanish (2), Ukrainian, Turkish, Yoruba, Tamil, Arabic (1 each)			Farsi (3), Arabic (3), Mandarin (2), Russian, French (1 each)		
Academic discipline	engineering (4), commerce (5), fine arts (1)			engineering (5), commerce (4), education (1)		
Student status	5 undergraduate, 5 graduate			4 undergraduate, 6 graduate		

Although the language of instruction at the university is English, the university population and the surrounding city are both multicultural and multilingual. Moreover, the university is in the province of Quebec, whose official language is French. Therefore, NN students had opportunities to hear and use various languages, including their L1s.

Because of different application requirements imposed by each academic program, some students had previously submitted standardized ESL test scores and were required to complete ESL courses, while others were not required to submit test scores or take ESL courses. A total of four final-semester students had submitted standardized scores prior to starting their studies and had completed at least two university ESL reading and writing courses, while six final-semester students had done neither. The ESL courses cover intermediate- and advanced-level reading and writing skills and vocabulary knowledge and use, with no emphasis on speaking or listening skills. None of the first- or final-semester students were taking or had taken university ESL speaking courses. Depending on the academic program, each student's final semester could fall in their second or third year of studies. In Quebec, almost all undergraduate university programs involve three rather than four years of full-time studies, while master's programs typically last two years. All the students in the final-semester group had studied between 24 and 32 months ($M = 30$) and were about to graduate from university.

2.1.2 Student raters

There were two groups of student raters: those studying human resource management (HRM) and those studying in other academic programs (non-HRM). The HRM group were 10 full-time undergraduate students from the university's commerce faculty ($M_{age} = 22.2$) who had received or were receiving instruction in human resource management. They had been recruited from courses in human resources management. These courses focused on staffing or had the staffing course as a pre-requisite. Research assistants visited the courses and posted notices inviting students from these courses to listen to job interview excerpts in return for financial compensation. Eligible and interested students contacted the research assistants. Most of the raters ($n = 6$) had completed one HRM course and were currently taking at least one other HRM course; nine raters had completed or were currently taking an HRM course focusing specifically on staffing. The course includes rigorous analysis of job descriptions, identification of essential competencies for various positions, and an extended simulation with students acting as interviewees or interviewers. Half of the raters were native speakers of English and half were non-native speakers. All assessed themselves on a 9-point scale of listening ability in English. One participant rated herself at 6; her data were discarded. The other 10 raters gave themselves 8 or 9 for English listening ability. In terms of exposure to non-native English, seven raters reported working in English with non-native speakers. All 10 raters reported their daily contact in English with non-native speakers at university as very frequent, and all named at least four different non-native accents they were familiar with (e.g., Hungarian, Vietnamese).

The non-HRM raters were 17 full-time students at the same university ($M_{age} = 25.4$) from different academic programs, primarily humanities and social science (11 undergraduate, 6 graduate). They were recruited through notices posted on the university campus inviting them to listen to L2 speech samples in return for financial compensation. All were native speakers of North American English. In terms of exposure to non-native English, nine reported working in English with non-native speakers. Five reported frequent or very frequent daily contact with non-native speakers at university and named four or five non-native accents they were familiar with. Twelve raters reported periodic or rare contact with non-native speakers and were familiar with French as well as one or two other non-native accents. There were many *potential* opportunities for non-HRM raters to be exposed to speakers of various L1s, since all were members of the multicultural university community. All raters from both groups reported normal hearing at the time of the study.

2.2 Task and procedure

The speaking task was a mock job interview. This task was selected because the speakers differed in their academic programs and standing, so they may have varied greatly in their knowledge of and experience with different academic speaking tasks. However, all speakers would have at least some knowledge of the purpose and overall structure of a job interview, so they would all be able to complete this task. Moreover, the task was authentic in that all speakers would likely engage in this type of speaking task during or after their studies.

Prior to the mock job interview, speakers filled out a consent form and a questionnaire in which they described their educational history, rated their abilities in speaking, listening, reading, and writing in English using a 9-point scale (1 = *very poor*, 9 = *very fluent*), and estimated the amount of time they typically heard or spoke English at school, at home, and socially. The speakers then read their assigned mock job posting (see Appendix). One mock job posting was for undergraduate and one for graduate students. Each job was designed so that students from a wide range of academic programs could potentially be hired. Each speaker read the relevant job posting, if necessary asking a research assistant questions about vocabulary. The speaker then had five minutes to prepare individually for the interview. After five minutes, any notes made by the speaker were kept by the research assistant, who gave the speaker another copy of the job posting and brought him or her to meet the interviewer. The interviewer was a female native speaker of English with over five years of experience teaching English to speakers of at least six different L1s. She also had experience interviewing genuine job applicants in English.

Each interview was recorded in a quiet room using an Olympus WS-321 digital recorder with a Sony ECM-T6 microphone. The interview resembled the content and format of a standard job interview. The interviewer first engaged the speaker in small talk for about a minute, then began asking the speaker about his or her qualifications and experience related to the job posting. Other standard interview questions, such as the speaker's strengths and weaknesses or future plans, were also asked. After answering any questions the speaker had about the mock position, the interviewer brought the interview to a close, then provided informal feedback to the speaker.

2.3 Data analysis

The interviews lasted a mean of 7.2 minutes (4.7–11.0). Speech samples of approximately one minute (0.6–1.0 min) were extracted as separate audio files from the

first four minutes of each recorded interview. Short audio samples were used in this study because pilot testing revealed that raters could complete their ratings within 30 seconds. Each speech sample began with the interviewer's question either about the speaker's suitability for the job or about the speaker's future plans, but the speaker did the majority of talking. When a speaker's complete answer to a question in the interview lasted more than one minute, the answer in the speech sample was limited to one complete idea (e.g., the speaker's ease with people, in response to a question about suitability for the position). These samples ended after a complete sentence, not mid-way through a sentence or phrase. Using speech editing software, all samples were normalized for intensity (loudness).

Each speech sample was evaluated by raters, tested individually in a quiet room. The raters first read the mock job postings and were informed that each speaker had interviewed for one of the two positions. The raters were instructed that they would hear excerpts of the job interviews consisting of speakers' answers to questions about qualifications or experience for the job or future career plans, and that they would use 9-point Likert scales to rate speakers on four aspects of speech. To clarify initial understanding of the constructs, the raters were told that accentedness (1 = *heavily accented*, 9 = *not accented at all*), comprehensibility (1 = *hard to understand*, 9 = *very easy to understand*), and fluency (1 = *not fluent at all*, 9 = *very fluent*) related to the raters' perceptions of each speaker's speech, not the content of the speaker's answer. They were then informed that communicative effectiveness on the job (1 = *poor communicative effectiveness on the job*, 9 = *excellent communicative effectiveness on the job*) targeted how effectively a speaker would communicate in a (non-specific) job requiring English. They were also told that this judgment required some prediction on their part which could go beyond judging only the speaker's speech.

The raters were randomly assigned to one of two randomly ordered sequences of speech samples, presented using Windows Media Player on a laptop computer with Sony stereo headphones. The raters clicked on the appropriate file and listened to the entire speech sample once without pausing, assigning four ratings to it during or after listening, and then moved on to listen to and rate the next speech sample. Before the main listening task, raters practiced by listening to and rating one speech sample from a speaker not included in the study.

3 Results

The first research question asked whether untrained university student raters would perceive differences in the L2 speech of NN students at the end of their

academic studies compared to NN students at the beginning of their studies. To determine if the raters' judgments were consistent, inter-rater reliability analyses were first carried out, separately for each rating in each rater group. The raters were consistent overall in their judgments (Cronbach $\alpha_{\text{non-HRM}} = .85-.95$, $\alpha_{\text{HRM}} = .91-.94$). Therefore, within each rater group, mean ratings for accentedness, comprehensibility, fluency, and communicative effectiveness were derived for each NN student by averaging across the raters' ratings. These mean ratings were then submitted to mixed two-way analyses of variance (ANOVAs) with speaker group (first-semester, final-semester) as the between-subjects factor and rater group (HRM, non-HRM) as the within-subjects factor. The ANOVAs (one for each of the four ratings) yielded no significant differences between speaker groups from the first and final semester, either as a main effect or an interaction, $F(1, 18) < 2.6$, $p > .12$, η^2_{partial} (effect size) $< .13$. Table 2 shows mean ratings for each rater group.

Table 2: Mean ratings of accentedness, comprehensibility, fluency, and communicative effectiveness for NN students in first and final semester of university studies (standard deviations in parentheses)

Raters	Accentedness		Comprehensibility		Fluency		Comm. effectiveness	
	First	Final	First	Final	First	Final	First	Final
Non-HRM ¹	3.6 (.9)	4.5 (1.6)	4.5 (1.2)	5.4 (1.4)	4.6 (1.2)	5.0 (1.4)	4.4 (1.3)	5.1 (1.3)
HRM ²	3.2 (.9)	4.2 (1.7)	5.2 (1.3)	5.9 (1.4)	4.8 (1.3)	5.3 (1.4)	4.7 (1.2)	5.3 (1.3)

¹ Non-HRM = raters not studying human resource management
² HRM = raters studying human resource management

The second research question asked whether raters studying human resource management evaluate NN students differently than do students from other academic programs. The ANOVAs described above yielded two significant main effects of rater group. More specifically, the HRM group was significantly harsher than the non-HRM group in rating speakers on accentedness (rating both groups of speakers as more accented), $F(1, 18) = 13.3$, $p = .002$, $\eta^2_{\text{partial}} = .43$. In contrast, the non-HRM group was significantly harsher than the HRM group in rating speakers on comprehensibility (rating both groups of speakers as harder to understand), $F(1, 18) = 13.0$, $p = .002$, $\eta^2_{\text{partial}} = .42$. There was no difference between rater groups in their ratings of fluency and communicative effectiveness. Because significant differences between the two rater groups were not qualified by significant interactions, Figure 1 displays these findings graphically for a combined group of 20 speakers.

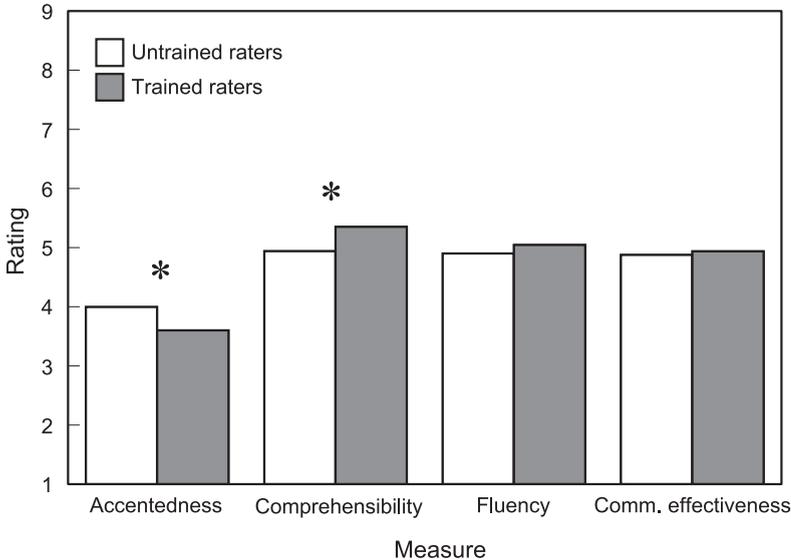


Fig. 1: Mean ratings for accentedness, comprehensibility, fluency, and communicative effectiveness by raters trained and untrained in human resource management. Asterisks designate statistically significant differences between the two rater groups.

4 Discussion

The goal of this study was to investigate the L2 speech of NN students in non-language-related programs and the rating behavior of untrained university student raters. NN students in their first and final semesters of study were recorded in mock job interviews. Interview excerpts were then presented to two rater groups, who rated the excerpts for accentedness, comprehensibility, fluency, and communicative effectiveness. Results showed that overall there was no difference in ratings for the first- and final-semester students, but that the two rater groups were different in their ratings of accentedness (HRM group harsher) and comprehensibility (non-HRM group harsher).

4.1 First versus final semester

The finding that there was no difference in ratings between first- and final-semester NN students is consistent with existing evidence from longitudinal research, namely, that some adult learners of English do not attain better

comprehensibility ratings over a period of two years living in an L2 environment (Derwing et al. 2008). However, evidence from cross-sectional research is mixed. Moyer (2011) found that accentedness ratings for international students at an English-medium university were related to students' length of residence in the university setting. In contrast, Piske et al. (2001) showed that the age of Italian immigrants when they first arrived in a predominantly English-speaking environment, not their length of residence in this environment, was an important predictor of the accentedness ratings of those immigrants. The two groups of speakers in this study were matched for age of arrival, so Piske et al.'s findings could explain the lack of difference between the first- and final-semester group in accentedness ratings, suggesting that age of arrival may be a stronger predictor of accentedness in adults, as compared to 24–32 months of “residence” at an English-medium university.

The lack of difference between the two groups' ratings in fluency and communicative effectiveness is somewhat surprising, especially with regard to fluency. One might expect that after years of study in English, with multiple opportunities to hear and produce the language orally, final-semester NN students would speak more smoothly, effortlessly, and with less hesitation than NN students who are just starting their studies. Lennon (1990) found that after living six months in an English environment, study-abroad learners of English received better fluency ratings than they had at the beginning of their stay. However, Derwing et al. (2008) found that some immigrant adult L2 learners did not receive better fluency ratings over a two-year period.

Another possible explanation for the lack of difference between the two speaker groups is that speakers' status as first- or final-semester students was less important than their use of or exposure to the L2, both in and outside the university setting. Some researchers have found that ratings of accentedness, comprehensibility, and fluency are correlated to self-reported measures of L2 use (Derwing et al. 2008; Moyer 2011), while others have found no links between listeners' ratings of L2 speech and speakers' reports of L2 use (Flege and Fletcher 1992). In order to determine whether, in this study, speakers' L2 exposure or use was important to the ratings they received, Pearson correlations were run between speakers' ratings and their self-reported daily hours of oral/aural L2 exposure or use at university, at home, and socially (range: 0–8 hours). No significant correlations were found between any of the exposure/use reports and any of the ratings. Implications of this lack of correlation are discussed below.

Finally, the non-significant results might be explained by two methodological aspects of the study. First, if differences between the two speaker groups were not large but more fine-grained, the use of untrained raters and Likert-type scales

to assess L2 speech may not have revealed these differences. However the same types of raters and scales were used in Derwing et al. (2008), where differences between speaker groups were found. Further research on using these types of raters and scales alongside discourse-level analyses of L2 speech (e.g., Kerekes 2007) may reveal the impact of using different approaches to measure and analyze L2 speech. Second, the size of the two speaker groups ($n = 20$) meant that each group's tendencies in L2 speaking ability may have been overshadowed by speakers' individual differences. The partial eta value of .13 (moderate effect size) suggests that a larger participant sample might have revealed differences between the two groups.

4.2 HRM raters

This study is the first to compare the behavior of untrained raters studying HRM and untrained raters in other academic programs. Results suggest that raters studying HRM do evaluate L2 speech somewhat differently than raters from other academic programs. The difference in exposure between rater groups may explain why HRM raters rated the NN students as more accented but more comprehensible than the non-HRM raters did. The HRM raters reported high levels of exposure to various non-native English speakers at work and in their daily lives at university, and reported familiarity with numerous non-native accents. In the faculty of commerce where the HRM raters studied, over half of the students in the faculty (54%) are non-native English speakers. In addition, a high proportion of faculty members are non-native speakers. In contrast, the non-HRM raters reported relatively less exposure to non-native speakers and accents, possibly because in the Faculty of Arts and Science, 36% of the students are non-native English speakers. In relative terms, then, the HRM raters, compared to the non-HRM raters, were exposed to more people speaking with varying levels of non-native English (from near-native English to English heavily influenced by speakers' native languages). This may have increased HRM raters' sensitivity to the range of variation which is possible (including near-native pronunciation), resulting in harsher accentedness ratings than the ratings given by the non-HRM group. This may also have made accented English more comprehensible to these raters because they had learned to adapt to non-native pronunciation (Bradlow and Bent 2008).

The HRM and non-HRM rater groups differed not only in their academic programs. Half of the raters in the HRM group were non-native English speakers, while no non-native English speakers were in the non-HRM group. However, the presence of non-native speakers in the HRM group appears to have played no role

in the group's rating behavior. The mean ratings of the native speaker raters and the non-native speaker raters were compared for all four aspects of speech, for the first- and then for the final-semester NN students. For each of the four aspects, the difference in ratings between native and non-native raters was half a point or less on a 9-point scale. Therefore, both the native and non-native raters in the trained group were rating speakers in very similar ways.

The lack of difference between rater groups for fluency and communicative effectiveness ratings suggests that the groups viewed the two aspects in similar ways. Unlike ratings of accentedness and comprehensibility, fluency ratings may be less influenced by raters' personal traits or experiences and more influenced by speakers' actual speaking rate and smoothness of delivery, which are not necessarily due to nativeness or non-nativeness. In fact, Derwing et al. (2009) found that the fluency ratings speakers received when speaking L2 English were related to the fluency ratings they received when speaking their native language. The similarity in ratings of communicative effectiveness may show that raters can largely agree when it comes to global impressions of speakers' ability to communicate in an L2. In future research, the use of verbal protocols during or after rating may help in identifying how raters' experience shapes their rating behavior (see Isaacs and Thomson 2013).

4.3 Implications for university programs

The first- and final-semester NN students received similar ratings on all four aspects of speech. This is despite the fact that the final-semester students had had considerably more opportunities at the university to hear authentic input, produce authentic output, and interact with English speakers. As noted above, the relatively small size of speaker groups may be one explanation for the lack of significant results. However, the results could indicate that when NN students follow their regular academic program, simply studying for a longer time in the L2 does not necessarily result in a greater degree of L2 speech development. The quantity of L2 exposure or use may be less important than the quality of L2 exposure or use. For example, Derwing et al. (2008) and Moyer (2011) noted that L2 speakers with more favorable accentedness or fluency ratings did not report more L2 exposure overall, compared to L2 speakers with less favorable ratings. However, speakers with better ratings reported more extended interactions using the L2.

NN students in English-medium universities have many opportunities to be exposed to and use English. However, simply having these opportunities may not be sufficient for the development of students' L2 speech (Crossling and Ward

2002). Therefore, many NN students could likely benefit from focused and sustained interventions which are aimed at increasing students' meaningful interactions in English, such as communicative listening and speaking courses or workshops, conversation groups, arranged pairings with other students or families, and mentorships. Results from the current study may suggest that in terms of development of L2 speech, NN students should focus not only on their academic performance but also on opportunities to have meaningful interactions in the L2, whether inside or outside the academic setting.

The finding that the HRM group found L2 speakers more comprehensible than the non-HRM group did may indicate that for proficient and native speakers of English, more exposure to L2 speakers promotes better comprehension of L2 speakers. It is therefore in the interests of both students and academic administrators to encourage interaction with L2 speakers in academic and other contexts. Students who are better able to understand L2 speakers will be better communicators both during and after their studies (see Krech Thomas 2004).

5 Conclusions and future directions

In this study, university students' ratings of the L2 speech of NN students was investigated through a cross-sectional design, with students in their first and final semesters being recorded. The obvious next step is a longitudinal study recording particular NN students at various points in their studies, which would give a fuller picture of individual students' L2 speech development and show how students' L2 learning is related to their academic progress. Another possible expansion of this study would be to analyze students' interactive speech, investigating how they work with interlocutors to "build" intelligibility and prevent and repair problems of understanding. To this end, it would be valuable to conduct stimulated recalls with each interlocutor in a given conversation in order to explore their thoughts and intentions while producing and interpreting L2 speech. Finally, the link between listeners' rating behavior and the nature of listeners' experiences should be further clarified.

The aim of this study was to explore the L2 speech development of NN university students at the beginning and end of their university careers. The lack of difference in ratings received by the two speaker groups suggests that NN students might benefit from extensive interventions to encourage sustained, meaningful interactions using the L2 over their university studies. NN students could then fully exploit the potential to develop their L2 speech while studying at university in their second language.

References

- Bradlow, Ann & Tessa Bent. 2008. Perceptual adaptation to non-native speech. *Cognition* 106(2). 707–729.
- Bridgeman, Brent, Donald Powers, Elizabeth Stone & Pamela Mollaun. 2012. TOEFL iBT test scores as indicators of oral communicative language proficiency. *Language Testing* 29(1). 91–108.
- Burger, Sandra & Marie Chrétien. 2001. The development of oral production in content-based second language courses at the University of Ottawa. *Canadian Modern Language Review* 58(1). 84–102.
- Carey, Michael, Robert Mannell & Peter Dunn. 2011. Does a rater's familiarity with a candidate's pronunciation affect the rating in oral proficiency interviews? *Language Testing* 28(2). 201–219.
- Cheng, Liying & Fox Janna. 2008. Towards a better understanding of academic acculturation: Second language students in Canadian universities. *Canadian Modern Language Review* 65(2). 307–333.
- Cheng, Liying, Johanne Myles & Andy Curtis. 2004. Targeting language support for non-native English-speaking graduate students at a Canadian university. *TESL Canada Journal* 21(2). 50–71.
- Crosling, Glenda & Ian Ward. 2002. Oral communication: The workplace needs and uses of business graduate employees. *English for Specific Purposes* 21(1). 41–57.
- Derwing, Tracey, Murray Munro & Ron Thomson. 2008. A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics* 29(3). 359–380.
- Derwing, Tracey, Murray Munro, Ron Thomson & Marian Rossiter. 2009. The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition* 31(4). 553–557.
- Derwing, Tracey, Murray Munro & Grace Wiebe. 1998. Evidence in favour of a broad framework for pronunciation instruction. *Language Learning* 48(3). 393–410.
- Derwing, Tracey, Marian Rossiter, Murray Munro & Ron Thomson. 2004. Second language fluency: Judgments on different tasks. *Language Learning* 54(4). 665–679.
- Díaz-Campos, Manuel. 2004. Context of learning in the acquisition of Spanish second language phonology. *Studies in Second Language Acquisition* 26(2). 249–273.
- Ferris, Dana. 1998. Students' views of academic aural/oral skills: A comparative needs analysis. *TESOL Quarterly* 32(2). 289–318.
- Flege, James, David Birdsong, Ellen Bialystok, Molly Mack, Hyekyung Sung & Kimiko Tsukuda. 2006. Degree of foreign accent in English sentences produced by Korean children and adults. *Journal of Phonetics* 34(4). 437–470.
- Flege, James & Kathryn Fletcher. 1992. Talker and listener effects on the perception of degree of foreign accent. *Journal of the Acoustical Society of America* 91(1). 370–389.
- Flege, James, Grace Yeni-Komshian & Serena Liu. 1999. Age constraints on second-language acquisition. *Journal of Memory and Language* 41(1). 78–104.
- Freed, Barbara. 1995. *Second language acquisition in a study-abroad context*. Amsterdam: John Benjamins.
- Freed, Barbara. 1998. An overview of issues and research in language learning in a study abroad setting. *Frontiers* 4, 31–60.

- Freed, Barbara, Norman Segalowitz & Dan Dewey. 2004. Context of learning and second language fluency in French. *Studies in Second Language Acquisition* 26(2). 275–301.
- Isaacs, Talia & Ron Thomson. 2013. Rater experience, rating scale length, and judgments of L2 pronunciation: Revisiting research conventions. *Language Assessment Quarterly* 10(2), 135–159.
- Kang, Okim & Dan Rubin. 2009. Reverse linguistic stereotyping: Measuring the effect of listener expectations on speech evaluation. *Journal of Language and Social Psychology* 28(4). 441–456.
- Kang, Okim, Rubin, Dan & Pickering, Lucy. 2010. Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English. *Modern Language Journal* 94(4). 554–566.
- Kennedy, Sara & Pavel Trofimovich. 2008. Intelligibility, comprehensibility, and accentedness of L2 speech: The role of listener experience and semantic context. *Canadian Modern Language Review* 64(3). 459–490.
- Kerekes, Julie. 2007. The co-construction of a gatekeeping encounter: An inventory of verbal actions [Special issue]. *Journal of Pragmatics* 39(11). 1942–1973.
- Krech Thomas, Helen. 2004. *Training strategies for improving listeners' comprehension of foreign-accented speech*. Boulder, CO: University of Colorado doctoral dissertation.
- Lasagabaster, David. 2008. Foreign language competence in content and language integrated courses. *Open Applied Linguistics Journal* 1. 31–42.
- Lee, Keun-Ho. 2006. *A traveler's tale: The experience of study in a foreign language*. Edmonton: University of Alberta doctoral dissertation.
- Lehtonen, Tuula & Sinikka Karjalainen. 2008. University graduates' workplace needs as perceived by employers. *System* 36(3). 492–503.
- Lennon, Paul. 1990. Investigating fluency in EFL: A quantitative approach. *Language Learning* 40(3). 387–417.
- Lyster, Roy. 2007. *Learning and teaching languages through content: A counterbalanced approach*. Amsterdam: John Benjamins.
- Moyer, Alene. 2011. An investigation of experience in L2 phonology. *Canadian Modern Language Review* 67(2). 191–216.
- Piske, Thorsten, Ian Mackay & James Flege. 2001. Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics* 29(2). 191–215.
- Rauto, Eeva. 2006. Changes in the interlanguage grammar of engineering students studying in an FL-medium program. [Special issue]. *Vaasan Yliopiston Julkaisuja* 132. 71–84. http://www.uva.fi/materiaali/pdf/isbn_952-476-149-1.pdf (accessed 4 September 2013).
- Ready, Doreen & Marjorie Wesche. 1992. An evaluation of the University of Ottawa's sheltered program: Language teaching strategies that work. In Robert Courchêne, Joan Glidden, Joy St. John & Christian Thérien (eds.), *Comprehension-based second language teaching / L'enseignement des langues secondes axé sur la compréhension*, 389–405. Ottawa: University of Ottawa Press.
- Segalowitz, Norman & Barbara Freed. 2004. Context, contact, and cognition in oral fluency acquisition. *Studies in Second Language Acquisition* 26(2). 173–199.
- Southwood, Helen & James Flege. 1999. Scaling foreign accent: Direct magnitude estimation versus interval scaling. *Clinical Linguistics and Phonetics* 13(5). 335–349.
- Zappa-Holman, Sandra. 2007. Academic presentations across post-secondary contexts: The discourse socialization of non-native English speakers. *Canadian Modern Language Review* 63(3). 455–485.

Appendix

Undergraduate Job Posting

Job Title: Survey Worker

This university takes pride in understanding and responding to the learning needs of its students. Your job will be to conduct surveys about teaching and learning at our university and to report on the results. The surveys will be created by the Office of Student Relations.

Duties

- Solicit possible survey participants, ask survey questions and record (write down) responses
- Calculate response totals, identify trends and patterns in responses
- Write brief (one-page) reports on survey results
- Give brief (five-minute) oral presentation on results to Office of Student Relations.

Qualifications

- Friendly, open manner with the public
- Self-motivated
- Good time-management, organizational, and analytical skills
- Good writing and presentation skills (summarizing ideas)
- Computer Skills-Word (required), Excel and statistical software (advantage)

Graduate Job Posting

Company Name: A1 Consulting

Job Title: Junior Project Coordinator

A1 Consulting is a dynamic firm which provides innovative services to businesses, technology firms, and educational organizations. We are searching for a Junior Project Coordinator to fill an entry-level position in our [academic program] team. The coordinator will track, manage, and report on several on-going consulting projects in [academic discipline]. Job tasks include: communicating with senior team members, setting up meetings, taking notes, and distributing documents. The performance of the Junior Project Coordinator will be reviewed after 3 months, with the possibility of promotion to a more senior position with greater responsibilities.

Qualifications

- Minimum Bachelor's degree in [academic program] (graduate-level studies are an advantage)

- Good Microsoft Office skills
- Excellent organizational and interpersonal skills
- Solid communication skills, both oral and written

Bionotes

Sara Kennedy is an assistant professor of applied linguistics in the Department of Education at Concordia University in Montreal, Canada. She conducts research on the teaching, learning, assessment and use of second language speech, particularly pronunciation, and on the role of language experience in the development of speaking ability.

Pavel Trofimovich is an associate professor of applied linguistics in the Department of Education at Concordia University. His research focuses on cognitive aspects of second language processing, second language phonology, sociolinguistic aspects of second language acquisition, and the teaching of second language pronunciation.

