Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Migyeong Jang April 3, 2017

Exploring Near-Nativeness in Second-Language English Speakers

by

Migyeong Jang

Dr. Marjorie Pak
Adviser

Program in Linguistics

Dr. Marjorie Pak

Adviser

Dr. Seth Goss

Committee Member

Dr. Jessica Barber

Committee Member

2017

Exploring Near-Nativeness in Second-Language English Speakers

By

Migyeong Jang

Dr. Marjorie Pak

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Program in Linguistics

2017

Abstract

Exploring Near-Nativeness in Second-Language English Speakers
By Migyeong Jang

Near-nativeness is a concept applicable to a highly proficient second-language (L2) speaker who is nearly indistinguishable from native (L1) speakers. In this study, L1 and L2 English speakers' performances in an oral production task and a Grammaticality Judgment Test (GJT) were examined, respectively aiming to assess proficiency in spoken communication and internal knowledge of English grammar. By gauging two possible close correlates of near-nativeness, the study strove to explore the concept and its implications. A positive yet imperfect significant correlation was found between the performances in the oral production task and the GJT, leading to the conclusion that GJT can be a useful measure of linguistic competence in near-native research. The results also showed that the mean scores of L1 speakers were significantly higher than that of L2 speakers in both the GJT and the oral production task. This finding is in favor of the idea that there is a meaningful difference in language use between L1 speakers and highly proficient L2 speakers. Acknowledging such differences can lead L2 learners to set more realistic and thus more appropriate goals of language learning.

Exploring Near-Nativeness in Second-Language English Speakers

By

Migyeong Jang

Dr. Marjorie Pak

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Program in Linguistics

2017

Acknowledgements

I would like to thank my committee members, Dr. Jessica Barber and Dr. Seth Goss, for their enthusiastic support and feedback. I would also like to thank my adviser, Dr. Marjorie Pak, for her unyielding patience, encouragement and guidance throughout the entire process.

Table of Contents

List of Figures

Introduction

The primary focus of this study is the notion of near-nativeness. To put it simply, distinguishing near-nativeness from nativeness is based on the conception that there is a limit in terms of ultimate attainment in second language (L2) acquisition. Under this premise, there would be meaningful differences between native speakers and "nativelike" L2 learners in their usage and/or knowledge of the language. For the purpose of this study, I am adopting the definition suggested by Abrahamsson and Hyltenstam (2008): "Near-native L2 proficiency is taken to mean apparent nativelikeness (i.e., levels of proficiency that cannot be distinguished from native levels in everyday spoken communication and only become apparent through detailed linguistic analyses)." (p. 484)

This definition of near-nativeness spotlights the **spoken communication** part of linguistic competence; however, competence in oral production of language does not necessarily have to be the sole measure of one's near-nativeness since there are so many different aspects of linguistic competence. One other factor that I suggest to be important in possibly achieving near-nativeness is one's **internalized grammatical knowledge**. In this study, I examine first language (L1) and L2 English speakers' performances in oral production and a Grammaticality Judgment Test, aiming to assess proficiency in spoken communication and internal knowledge of English grammar, respectively. By gauging two possible close correlates of near-nativeness, I aim to explore the concept and its implications.

Grammaticality Judgment Test (GJT) is a method of assessing one's knowledge of grammar in the target language, which is often used in the field of Second Language Acquisition (SLA). In a GJT, the subjects are given multiple sentences in the language of interest and asked to decide whether each sentence is grammatical or not. It is relatively quick and easy to conduct.

More importantly, its results can be conveniently scored and quantified unlike some other methods used in SLA – for instance, a natural production task. There have been several studies that looked at their validity and reliability in the context of SLA.

Mandell (1999) investigated if there was a correlation between the score on a GJT and the score on a dehydrated sentence test regarding certain syntactic properties in Spanish. Dehydrated sentence (DS) test, or a slash-sentence test, is a task which asks the subject to build a grammatical sentence by placing the provided items in the right order and converting them into the right forms. It was chosen as a comparative measure to the GJT since it is commonly used in L2 teaching as a way to assess the learner's linguistic knowledge. While both are written assessments, GJT and DS test differ meaningfully in that DS test has a requirement for production on the subject's part. The researcher found that the correlations between the GJT and the DS test were consistent as well as cross-sectional. And based on these results, he concluded that GJTs are, indeed, a reliable method of measuring the linguistic competence of L2 users. This study has practical implications especially important in the field of SLA in that it found correspondence in two often-used written measures of linguistic competence.

Leow (1996) looked at the validity of GJTs in a more specific context, which was L2 development. This study examined the validity of GJTs by investigating whether there was a correlation between the learners' performances on a GJT and those on a production task. Since the researcher tried to address this question specifically in relation to L2 development, the same tasks were conducted twice – once after 6 hours of L2 Spanish instruction, and again after 35 hours of instruction. GJT scores and production task scores significantly correlated in both stages of L2 development, leading to the author's claim that GJTs accurately reflect the development patterns of L2 learners, and thus can be deemed as a valid method of measuring L2 knowledge.

While Leow (1996) has a study design similar to the current research, the key difference between the two studies lies in the proficiency level of the subjects. The populations of interest are almost polar opposites from each other in that Leow (1996) looked at beginner-level L2 learners while the current study focuses on highly proficient L2 users who are potentially indistinguishable from native speakers.

On the other hand, a 2012 article by Tabatabaei and Dehghani is an example of a study which concluded that the GJT is an unreliable reflection of one's grammatical knowledge. In this study, instead of comparing the GJT to another method of linguistic assessment, they focused on the test-retest reliability and internal consistency of the GJT itself. As opposed to Mandell (1999) and Leow (1996), they argued that performance on GJTs can often fluctuate regardless of actual linguistic competence. Conclusions from this study cannot be easily generalized, however, since as made clear by the researchers, the findings only apply to the specific GJT that was used in the study itself. The fact that one specific GJT led to unreliable results does not necessarily imply that the GJT as a whole is unreliable.

As such, there is yet to be a consensus on the validity and efficacy of the GJT. Furthermore, there has not been an abundance of comprehensive research that investigated GJTs in the specific context of near-nativeness. Adopting the right method is an essential step for further investigating a concept that is not yet clearly established (i.e., near-nativeness). Thus, examining whether GJTs are the right tool for this task seems to be a meaningful undertaking. I aim to compare the results of a GJT and an oral production task in potentially near-native L2 users of English. By comparing these two methods, I hope to find out if using GJTs is an appropriate method of measuring near-nativeness. The main question here is whether their results correlate with each other. I hypothesize that there will be a positive correlation between

GJT scores and production task scores. The two tasks will measure the same thing – near-nativeness – in general, but the results are unlikely to exactly match each other all the time.

The potential implications of near-nativeness research are not only confined to the field of SLA. Coppieters, who defines a near-native speaker as "native-like non-native speaker," (Coppieters, 1987, p. 544) claims that exploring possible differences in language use between native and near-native speakers poses us a more fundamental question about language: "Does a language impose a grammar on its speakers? Is the internal grammar of a language, as reflected by a speaker's intuitions about sentences of that language, a largely stable result of the interaction of (a) some cognitive learning mechanism and (b) the linguistic input provided by a language community?" (Coppieters, 1987, p. 544) As such, examining validity of near-nativeness as a concept can guide us back to bigger questions in the field of linguistics.

Method

Participants

The sample consisted of 50 currently enrolled undergraduate students at Emory University in Atlanta, Georgia. The age range of the participants was from 18 to 25. Thirty-five participants were L1 speakers of American English. Fifteen participants were L2 speakers of English, and their L1 was controlled to Korean in order to minimize any unexpected environmental and cognitive variation in their use of English. As undergraduate students, the group of L2 English speakers were expected to have little to no difficulty in conversational and academic use of the English language.

Participants were recruited using convenience sampling. For recruitment, an electronic flyer and announcements that briefly explained the study and procedures were posted on Facebook. Potential participants were asked to contact the researcher via e-mail or telephone to individually schedule their study visits.

Materials

Demographics questionnaire. At the beginning of the study visits, all 50 participants were asked to complete a demographic questionnaire designed for the study. The questionnaire included items about age, sex, ethnicity, L1, and the percentage of their daily interaction conducted in English. For those who answered that their L1 was not English, two additional items were asked about their background in English: time resided in an English-speaking country and the age at which they started learning English. (Appendix A)

Oral production task: interview questions. All 15 L2 speakers of English and 5 control L1 speakers participated in an oral production task. The interview was conducted one-on-one with the researcher, and each interview session was recorded. The oral production task was designed to assess everyday spoken communication of the speaker. The first two questions were counting and listing, which primarily focused on any possible phonetic differences. The third question was to elicit spontaneous speech which would lead to the use of a variety of vocabulary items and expressions. For the last part of the interview, each speaker was given an identical paragraph to read out loud. Each interview took five to seven minutes. (Appendix B)

Questionnaire A: Grammaticality judgment test (GJT). All 15 L2 speakers of English and 5 control L1 speakers completed questionnaire A, which was designed to assess grammatical knowledge of the speaker. Questionnaire A consisted of two parts: traditional grammaticality

judgment task (Part I: 40 questions) and fill-in-the-blank style production task (Part II: 10 questions). All questions were weighted equally. All participants conducted the task on paper under monitoring. Each participant was given 10 minutes to finish the questionnaire, and all participants finished before the given time.

To measure more nuanced grammatical knowledge of the English language, especially any possible difference in grammatical knowledge between L1 and L2 speakers, I chose the definite article *the* as the key variable. The use of the definite article is interesting in that even highly fluent L2 English speakers often make mistakes concerning it (Edmunds, 2013). While there are rules regarding its usage, exactly what properties lead to "definiteness" can often be hard to determine (Abbot, 2014). Hence, it seems to be an aspect of grammatical knowledge that comes rather instinctively to L1 speakers. That it confuses not only beginner-level but also proficient L2 English speakers and that it cannot simply be acquired through memorization make *the* an interesting grammar feature to look at. Thus, the questionnaire focused on English articles – including but not limited to the definite article. It is noteworthy that neither Korean, the controlled L1 of the current study, nor Mandarin Chinese, the controlled L1 of Edmunds (2013), is a language that uses a definite article in a sense that the English language uses 'the.'

In Part I, all question items contained either a correct or an incorrect usage of *a, an,* or *the*. All participants were instructed to read each sentence and decide if it sounded grammatical or not. Further, they were instructed to fix any grammatical error they found. If they thought there were multiple errors in a single item, they had to fix what they considered to be the biggest error. For the purpose of scoring, the corresponding score was subtracted if they correctly marked an item as ungrammatical but for a wrong reason. Out of 40 items total, 16 items (40%) were grammatical sentences, and 24 items (60%) were sentences with an erroneous use of

articles. Out of the 24 sentences with a grammatical error, 12 sentences (50%) involved overuse errors (using definite or indefinite articles when unnecessary or inappropriate), and another 12 sentences (50%) involved underuse errors (omitting definite or indefinite articles when necessary). Sentence length (counted by number of words) averaged 10 words across all 40 items: 9.5 words in grammatical items, 10.4 words in overuse error items, and 10.1 words in underuse error items. Average word length (counted by number of syllables) was 1.5 syllables per word in grammatical items and 1.4 syllables per word in both overuse and underuse error items. For the error items, the placement of the error was balanced as well. For each error type respectively, there were 6 items (50%) with the error in the first half of the sentence and another 6 (50%) with the error in the latter half of the sentence.

In Part II, participants had to complete a paragraph by filling in the blanks using *a, an, the,* or no article. Out of 10 blanks, 4 items (40%) were to be an indefinite article *a*, 4 items (40%) to be *the*, and the remaining 2 (20%) to be no article. (Appendix C)

Questionnaire B: Perception task. 30 L1 speakers of English participated in the perception task. The perception task involved filling out a written questionnaire while listening to audio materials extracted from the oral production task. Among the four questions asked in the interview, the first question (counting) and a portion of the last question (passage reading) were selected. Part I of the questionnaire was for counting, and part II for passage reading. For each part, all 20 speakers – 5 L1 controls and 15 L2 speakers – were randomized in order.

Both Part I and II had the same instructions: to listen to each recording and indicate how nativelike the speaker sounds. A 4-point Likert scale was used, with written descriptions for both endpoints: 1 = “non-native; English is not their first language – learned English as a second or

third language," 4 = "native; grew up speaking English as their first or only language." Each part consisted of 20 items, one per each speaker of the recordings. (Appendix D)

Study Design and Procedure

While the oral production task aimed to assess one's everyday spoken communication, the GJT was supposed to measure one's grammatical knowledge. Based on the idea that these two would measure at least partially different aspects of one's language, it was hypothesized that there would be a generally positive yet imperfect correlation between the performances in the oral production task and the GJT.

The perception task conducted by L1 speakers was chosen to be the method of assessing performance in the oral production task, not only because oral performances are nearly impossible to be quantified by a single scorer, but also because it is, indeed, the best way to see if the L2 speakers can pass as L1 speakers because of their high proficiency. In the end, if a large number of L1 speakers mistakenly believe a large number of L2 speakers to be native, that could potentially lead us to the idea that there might not be a significant difference in language use between L1 speakers and very proficient L2 speakers.

In addition to probable differences between performances in the two measures within an individual, differences in performances across individuals were expected. Would there be a significant difference between L1 and L2 speakers? Among L2 users, what factors (e.g., age of onset) would show correlation with performance?

All data collection procedures were conducted on Emory's main campus, in person. A written form of informed consent was obtained before any data collection. Participants were asked to read the informed consent statement and indicate their consent by signing the copy.

After filling out the demographics questionnaire, 5 L1 controls and 15 L2 speakers proceeded to complete the oral production task and then questionnaire A. 30 L1 speakers completed the perception task while listening to the audio materials extracted from the oral production task.

Results

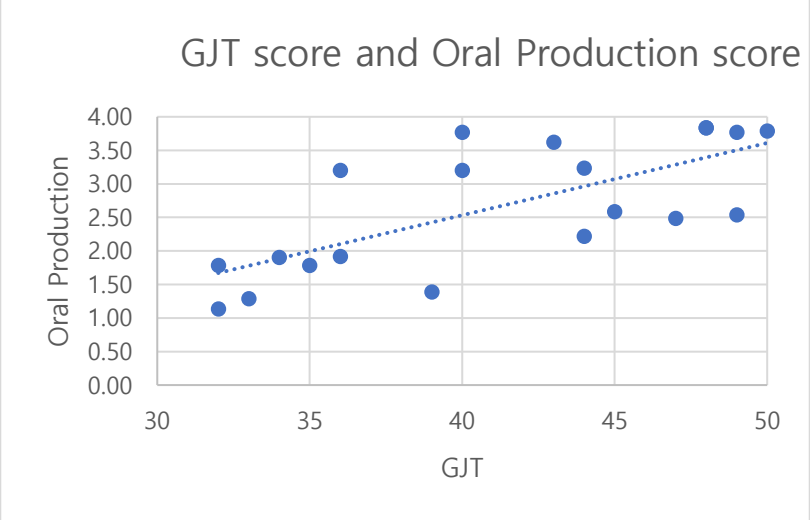
Participant	Oral Production Task			GJT			Demographics	
#	Part 1	Part 2	Total	Part 1	Part 2	Total	Residence	Initiation
2	3.67	4	3.83	38	10	48	N/A	N/A
4	3.73	3.93	3.83	38	10	48	N/A	N/A
1	3.67	3.9	3.78	40	10	50	N/A	N/A
3	3.8	3.73	3.77	30	10	40	N/A	N/A
5	3.6	3.93	3.77	39	10	49	N/A	N/A
20	**3.63**	**3.6**	**3.62**	**33**	**10**	**43**	**14**	**1**
8	3.07	3.4	3.23	35	9	44	7	6
7	2.77	3.63	3.20	31	9	40	13	6
10	3.4	3	3.20	28	8	36	3	12
16	2.9	2.27	2.58	36	9	45	4	10
12	**2.73**	**2.33**	**2.53**	**39**	**10**	**49**	**4**	**8**
6	**2.27**	**2.7**	**2.48**	**38**	**9**	**47**	**10**	**6**
15	2.77	1.67	2.22	34	10	44	12	6
13	1.87	1.97	1.92	28	8	36	5	9
14	2.4	1.4	1.90	25	9	34	8	9
11	2.17	1.4	1.78	25	7	32	6	10
17	2.1	1.47	1.78	29	6	35	2.3	7
9	1.57	1.2	1.38	29	10	39	10	13
18	1.33	1.23	1.28	25	8	33	6	10
19	1.13	1.13	1.13	25	7	32	3	7

< Table 1 >

Table 1 is a summary of scores from both the oral production task (i.e. mean scores speakers got for their performance in the oral production task) and the GJT. Participants are ranked by their mean scores from the oral production task. Participants shaded in gray (#1-5) are

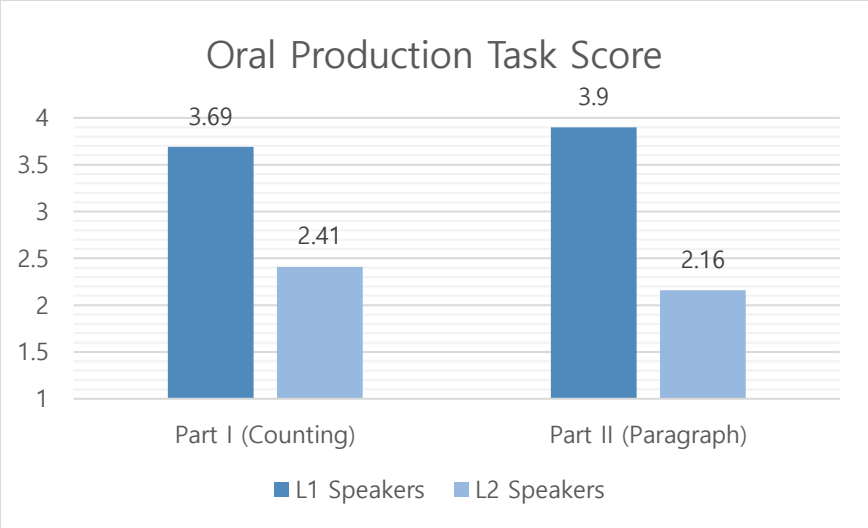
L1 controls. The score range for oral production task is from 1 to 4, 1 being considered non-native and 4 being considered native by the listeners. The scores were rounded up to two decimal points. The score range for the GJT is from 0 to 50 (0-40 for Part 1, 0-10 for Part 2).

The hypothesis that there would be a positive correlation between the performances in the oral production task and the GJT was supported. Graph 1 shows the significant correlation (r=0.71, p<.001) that was found between the GJT scores and perception task scores.

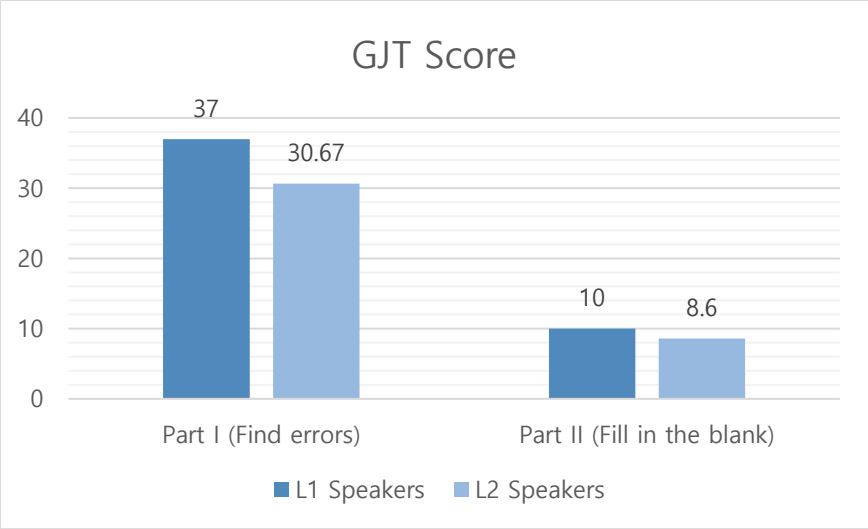


< Graph 1 >

When examined by 2-tailed independent t-test, the mean scores of L1 speakers were significantly higher than that of L2 speakers in both GJT (t=2.78, p<.05) and the oral production task (t=4.27, p<.05). This finding is in line with the idea that there is a meaningful difference in language use between L1 speakers and highly proficient L2 speakers. In the GJT, the gap in performance between L1 and L2 speakers was bigger in Part 1 (traditional GJT items) than in Part 2 (fill-in-the-blank items). In the oral production task, the score difference between L1 and L2 speakers was bigger in Part 2 (sentence reading) than in Part 1 (counting).



< Graph 2 >



< Graph 3 >

A significant positive correlation (r=0.89, p<.001) was found between performances in Part I and Part II in the oral production task. In the GJT, scores from Part 1 and Part 2 showed a significant positive correlation (r=0.67, p=.001) as well. These correlations can be interpreted to argue that different parts of each task are measuring mostly the same aspects of language knowledge, as they were intended to.

For L2 speakers (participants #6-20), "residence" column refers to the total number of years they have been residing in English-speaking countries, and "initiation" column refers to the age they started learning English. The length of residence positively correlated with the scores in the oral production task (r=0.38) and the GJT (r=0.35), even though the correlation was not statistically significant (p>.05). The age of initiation, on the other hand, negatively correlated with score in oral production (r=-0.49) and GJT (r=-0.38); in other words, the younger they were when they started learning English, the better they performed in both tasks. The correlation was not statistically significant in this case either (p>.05).

It is noteworthy that speakers who got the five highest mean scores in the oral production task – in other words, speakers who were considered to sound most "nativelike" by L1 speakers – were, indeed, L1 speakers. Participants in bold (#20, #12, #6) are L2 speakers who performed exceptionally well in either one of the tasks. However, none of the L2 speakers performed extraordinarily well in both tasks. Participant #20 scored sixth highest in the oral production task, but his/her performance in GJT was not extraordinary. Her age of onset for English acquisition was 1, which was by far the youngest among the L2 speakers. Conversely, participants #6 and #12 scored very high on GJT but not on the oral production task.

Participant #3 was an exception in that his/her GJT score was significantly lower than that of all the other L1 speakers as well as several L2 speakers. One possible explanation for his/her low GJT score is that he/she was not very thorough when completing the task, based on the observation that 9 out of 10 errors were made in underuse items and all in Part 1. In other words, it is possible that the individual glanced through the sentences, filling in the absent articles. Below are examples of underuse items:

**Marilyn is only person I can trust with anything.*
(Marilyn is <u>the</u> only person I can trust with anything.)

**Last night, Anne and I finally went to restaurant that her sister loves.*
(Last night, Anne and I finally went to <u>the</u> restaurant that her sister loves.)

Such pattern of errors made by participant #3 is especially interesting when looking at the errors L2 speakers made in Part 1. With L2 speakers, the biggest number of errors were made with overuse items. For each question that contained an overuse error, an average of 4.5 participants put down an incorrect answer. With underuse items the average was lower at 3.75, and with correct items the average was the lowest at 2.56. In this specific GJT, overuse error translates to inserting *the* where it is not needed. Below are examples of overuse items:

**We are planning to go on a romantic date on the Valentine's Day.*
(We are planning to go on a romantic date on Valentine's Day.)

**Getting enough sleep is a big problem for the college students.*
(Getting enough sleep is a big problem for college students.)

L2 English speakers showing a pattern of overusing the definite article aligns with the results reported by Edmunds (2013) in her corpus study of Chinese ESL learners. Additionally, that those participants performed better on errorless sentences than on sentences with a grammatical error parallels the results demonstrated in Murphy (1997) and Gutiérrez (2013) – people are better at judging a grammatical sentence as grammatical than at judging an ungrammatical sentence as ungrammatical.

Discussion

Like any other method of measuring linguistic ability of an individual, the GJT is far from perfect. Nevertheless, it can still be a powerful window to peer into one's grammatical knowledge, especially considering its accessibility and convenience. Using the GJT in conjunction with other research methods can be even more advantageous; for instance, a 1989 study by Bley-Vroman and Masterson used reaction time as a supplement to the GJT and obtained consistent results.

From the positive correlation between the performances in the oral production task and the GJT, it can be deduced that the GJT is a useful measure of linguistic competence in near-nativeness research. Since the oral production task assesses the perception of spoken communication – which Abrahamsson and Hyltenstam (2008) argue to be the key factor of near-nativeness – and the GJT significantly correlates with that task, performance in the GJT should be considered a close correlate of near-nativeness.

Statistically significant score differences in both tasks between L1 and L2 speakers can help strengthen the claim that there is, indeed, a meaningful difference in language use between L1 speakers and highly proficient L2 speakers. This finding is contradictory to the results from White and Genesee (1996), which found no significant difference in performance between near-native and native speakers. However, as mentioned by the authors themselves, it is likely that different subject selection processes can lead to different results. Unlike this study, which simply used convenience sampling within the population of interest, White and Genesee went through an additional filtering process in which they labeled only the highest performing individuals as near-native. This discrepancy between the two studies suggests possible ramifications different sampling processes can bring.

In most L2 speakers, a noticeable gap between the performances in oral production and grammatical knowledge was observed. Even the participants who performed on the level of a native speaker in one of the tasks scored notably lower on the other task. These results can be tied back to Han (2006), a study of fossilization in L2 acquisition. In this article, the author suggests the possibility that there might be several local endpoints instead of one global endpoint when it comes to ultimate attainment. In other words, it is entirely possible for an individual to achieve nativelike ability in one aspect of language and but not all of them. This poses an interesting question regarding the definition and criteria of near-nativeness: would some facets of linguistic competence matter more than others? If so, which ones?

This study can have several practical implications for L2 instruction. First, the aforementioned gap between different task performances highlights the importance of L2 education that is balanced across different facets of language. For instance, solely focusing on colloquial communication might lead to lacking knowledge in terms of L2 grammar. Second, focusing on the pattern of errors that L2 speakers make – in this study, overuse errors of the definite article – can pinpoint the source of discrepancies in understanding. Finally, and on a more fundamental level, the observed difference in linguistic competence between L1 and L2 speakers can lead L2 learners to set more realistic and thus more appropriate goals of language learning. Birdsong (2005) points out how detrimental it can be to set nativelikeness as a standard for all L2 learners without acknowledging reasonable limitations. As such, proficiency criteria for L2 learners should recognize and adapt to their limitations rather than replicate L1 criteria.

Future research could utilize bigger sample sizes, attending to the differences the sampling process can make. For instance, recruiting participants who consider themselves near-native might lead to a group of exceptionally competent individuals who are very confident in

their L2 performance. Looking at different modalities can be interesting as well. Murphy (1997) reported that participants in the aural condition performed less accurately and more slowly than those in the visual condition during a GJT. Different proficiency level of the subjects can lead to different results as well.

References

Abbot, B. (2014). The indefiniteness of definiteness. In Gamerschlag, T., Gerland, D., Osswald, R., & Petersen, W (Eds.), *Frames and Concept Types* (pp. 323-341). Springer International Publishing.

Abrahamsson, N., & Hyltenstam, K. (2008). The robustness of aptitude effects in near-native second language acquisition. *Studies in second language acquisition, 30*(04), 481-509.

Bijvoet, E. (2002). Near nativeness and stylistic lexical competence in Swedish of first and second generation Finnish immigrants to Sweden. *International Journal of Bilingualism, 6*(1), 39-51.

Birdsong, D. (2005). Nativelikeness and non-nativelikeness in L2A research. *International Review of Applied Linguistics in Language Teaching, 43*(4), 319-328.

Bley-Vroman, R., & Masterson, D. (1989). Reaction time as a supplement to grammaticality judgments in the investigation of second language learners' competence. *University of Hawai'i Working Papers in ESL, 8*(2), 207-237.

Coppieters, R. (1987). Competence differences between native and near-native speakers. *Language*, 544-573.

Edmunds, K. (2013). *Chinese ESL learners' overuse of the definite article: a corpus study*. Retrieved from Emory Electronic Theses and Dissertations Repository. (http://pid.emory.edu/ark:/25593/d7xt0)

Gutiérrez, X. (2013). The construct validity of grammaticality judgment tests as measures of implicit and explicit knowledge. *Studies in Second Language Acquisition, 35*(3), 423-449.

Han, Z. (2006). Fossilization: Can grammaticality judgment be a reliable source of evidence. *Studies of fossilization in second language acquisition*, 56-82.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive psychology, 21*(1), 60-99.

Leow, R. (1996). Grammaticality judgment tasks and second-language development. In Alatis, J., Straehle, C., Ronkin, M., & Gallenberger, B (Eds.), *Georgetown University round table on languages and linguistics* (pp. 26-39). Washington, D.C: Georgetown University Press.

Mandell, P. B. (1999). On the reliability of grammaticality judgement tests in second language acquisition research. *Second Language Research, 15*(1), 73-99.

Murphy, V. A. (1997). The effect of modality on a grammaticality judgement task. *Second Language Research, 13*(1), 34-65.

Tabatabaei, O., & Dehghani, M. (2012). Assessing the reliability of grammaticality judgment tests. *Procedia-Social and Behavioral Sciences, 31,* 173-182.

White, L., & Genesee, F. (1996). How native is near-native? The issue of ultimate attainment in adult second language acquisition. *Second language research, 12*(3), 233-265.

Appendix A

Demographics Questionnaire

1. What is your age?

_____

2. How do you identify yourself?

A. Female
B. Male
C. Other _____

3. What is your ethnicity? (Circle all that apply)

A. Caucasian / White
B. Black / African American / African
C. Asian / Asian American / Pacific Islander
D. Hispanic / Latino
E. Indigenous / Native American Descent
F. Other _____

4. What is your first language?

_____

4a. If your first language isn't English:

For how long have you been living in an English speaking country?

If there have been multiple time periods, please write down the total amount of time.

_____ (years)

4b. If your first language isn't English:

How old were you when you first started learning English?

_____ (years)

5. What percentage of your daily interaction (including written language) is conducted in English?

A. 100%
B. 80 - 99%
C. 60 - 79%
D. 40 - 59%
E. Less than 40%

Appendix B

Interview Questions

1. Tell me the days of the week.

2. Count from one to ten.

3. Think about your favorite movie or book. Briefly summarize the story for me.

4. Read the paragraph out loud.

“We had some sad news last night. A man named Alan Brown, one of our neighbors who used to own an art gallery in Old City, suffered a massive stroke while he was swimming laps at the gym and had to be rushed to the hospital in an ambulance. He passed away later that night. He was a very sweet old man who swam at the Y every day. He wore big, horn-rimmed glasses and had a very charming manner of speaking, and he loved to tell jokes and make people laugh. His passing has upset everyone.”

Appendix C

Questionnaire A: Grammaticality Judgment Task

Part I. Please read each sentence and decide if it sounds okay to you (mark ✓ or X). If you think the sentence contains a grammatical error, please fix the error. If you think there are multiple errors, please fix what you consider the biggest problem.

1. I accidentally punched Joshua in the face.
2. I bet she's from the South. Listen to how she talks!
3. Getting enough sleep is a big problem for the college students.
4. 40% of the Americans ages 10-20 say that they watch 3 hours of TV per day.
5. This is best steak I have ever had in my entire life.
6. Jake is going to Philippines this summer.
7. Saving the water is important for the environment.
8. Have you ever thought about adopting a cat?
9. I went to the bed really early last night.
10. My hobby is watching the movies.
11. Veronica's dad had a heart attack. I hope he feels better soon.
12. There are 159 counties in the state of Georgia.
13. What are your plans for Thanksgiving this year?
14. I am the Christian. I go to church every Sunday.
15. It took me a whole day to read this chapter.
16. I'm feeling awful. I got the flu.
17. I want to learn to play the tennis.
18. Have you had the lunch yet? If not, let's go grab something to eat!
19. Last night, Anne and I finally went to restaurant that her sister loves.
20. In Korea, most children go to the school five days a week.
21. Jack had a little difficulty solving calculus problems.
22. The GPS says that we are heading north.
23. In 1920s, the U.S. government started enforcing a strict immigration policy.
24. The French is a difficult language to learn.
25. The brunch place that John and I went to yesterday was amazing.

26. We are planning to go on a romantic date on the Valentine’s Day.
27. This morning, I had to take a taxi to work because my car wouldn’t start.
28. Where did the time go? It’s already last day of fall break.
29. Jen has been conducting research about reproductive rights for years.
30. Could you pass me salt?
31. Have you met Sophie? She’s sweetest kid ever.
32. I want to travel across Sahara Desert one day.
33. Don’t touch the pizza in the fridge. It’s mine!
34. Today is the first day of summer vacation.
35. I always wanted to learn how to play the guitar.
36. Did you read book that I gave you?
37. Marilyn is only person I can trust with anything.
38. Do you know name of his puppy?
39. Some children like to read the books for fun.
40. I got internship that I told you about!

Part II. Please fill in the blanks. You can use **a**, **an**, **the**, or **no article**.

Ice cream is _____ sweetened frozen food typically eaten as _____ snack or dessert. _____ meaning of _____ phrase "ice cream" varies from one country to another. In some countries, such as _____ United States, _____ phrase "ice cream" applies only to _____ specific variety. Ice cream may be served in _____ dishes, for eating with _____ spoon, or in _____ cones, which are licked.

Appendix D

Questionnaire B: Perception Task

Please listen to the recordings and indicate how nativelike the speaker sounds in your opinion (circle one of the four choices).

PART I

	1 (non-native; English is not their first language - learned English as a second or third language)	**2**	**3**	**4 (native; grew up speaking English as their first or only language)**
Speaker 1	1 (non-native)	2	3	4 (native)
Speaker 2	1 (non-native)	2	3	4 (native)
Speaker 3	1 (non-native)	2	3	4 (native)
Speaker 4	1 (non-native)	2	3	4 (native)
Speaker 5	1 (non-native)	2	3	4 (native)
Speaker 6	1 (non-native)	2	3	4 (native)
Speaker 7	1 (non-native)	2	3	4 (native)
Speaker 8	1 (non-native)	2	3	4 (native)
Speaker 9	1 (non-native)	2	3	4 (native)
Speaker 10	1 (non-native)	2	3	4 (native)
Speaker 11	1 (non-native)	2	3	4 (native)
Speaker 12	1 (non-native)	2	3	4 (native)
Speaker 13	1 (non-native)	2	3	4 (native)
Speaker 14	1 (non-native)	2	3	4 (native)
Speaker 15	1 (non-native)	2	3	4 (native)
Speaker 16	1 (non-native)	2	3	4 (native)
Speaker 17	1 (non-native)	2	3	4 (native)
Speaker 18	1 (non-native)	2	3	4 (native)
Speaker 19	1 (non-native)	2	3	4 (native)
Speaker 20	1 (non-native)	2	3	4 (native)

PART II

Text: “A man named Alan Brown, one of our neighbors who used to own an art gallery in Old City, suffered a massive stroke while he was swimming laps at the gym and had to be rushed to the hospital in an ambulance.”

	1 (non-native; English is not their first language - learned English as a second or third language)	**2**	**3**	**4 (native; grew up speaking English as their first or only language)**
Speaker 1	1 (non-native)	2	3	4 (native)
Speaker 2	1 (non-native)	2	3	4 (native)
Speaker 3	1 (non-native)	2	3	4 (native)
Speaker 4	1 (non-native)	2	3	4 (native)
Speaker 5	1 (non-native)	2	3	4 (native)
Speaker 6	1 (non-native)	2	3	4 (native)
Speaker 7	1 (non-native)	2	3	4 (native)
Speaker 8	1 (non-native)	2	3	4 (native)
Speaker 9	1 (non-native)	2	3	4 (native)
Speaker 10	1 (non-native)	2	3	4 (native)
Speaker 11	1 (non-native)	2	3	4 (native)
Speaker 12	1 (non-native)	2	3	4 (native)
Speaker 13	1 (non-native)	2	3	4 (native)
Speaker 14	1 (non-native)	2	3	4 (native)
Speaker 15	1 (non-native)	2	3	4 (native)
Speaker 16	1 (non-native)	2	3	4 (native)
Speaker 17	1 (non-native)	2	3	4 (native)
Speaker 18	1 (non-native)	2	3	4 (native)
Speaker 19	1 (non-native)	2	3	4 (native)
Speaker 20	1 (non-native)	2	3	4 (native)

Appendix E

Questionnaire A: Grammaticality Judgment Task – Answer Keys

Part I. Grammaticality Judgment task (40 items)

16 Correct items (40%)

- Have you ever thought about adopting a cat?

- Jen has been conducting research about reproductive rights for years.

- The brunch place that John and I went to yesterday was amazing.

- This morning, I had to take a taxi to work because my car wouldn't start.

- Veronica's dad had a heart attack. I hope he feels better soon.

- It took me a whole day to read this chapter.

- I'm feeling awful. I got the flu.

- The GPS says that we are heading north.

- Jack had a little difficulty solving calculus problems.

- I always wanted to learn how to play the guitar.

- Don't touch the pizza in the fridge. It's mine!

- I bet she's from the South. Listen to how she talks!

- I accidentally punched Joshua in the face.

- Today is the first day of summer vacation.

- There are 159 counties in the state of Georgia.

- What are your plans for Thanksgiving this year?

12 Overuse/Insertion error (30%)

- 40% of (the) Americans ages 10-20 say that they watch 3 hours of TV per day.

- I am (the) Christian. I go to church every Sunday.

- I went to (the) bed really early last night.

- Saving (the) water is important for the environment.

- Have you had (the) lunch yet? If not, let's go grab something to eat!

- (The) French is a difficult language to learn.

- Getting enough sleep is a big problem for (the) college students.

- My hobby is watching (the) movies.

- I want to learn to play (the) tennis.

- We are planning to go on a romantic date on (the) Valentine's Day.

- Some children like to read (the) books for fun.

- In Korea, most children go to (the) school five days a week.

12 Underuse/Deletion error (30%)

- In 1920s, the U.S. government started enforcing a strict immigration policy.

- This is best steak I have ever had in my entire life.

- Marilyn is only person I can trust with anything.

- Do you know name of his puppy?

- I got internship that I told you about!

- Did you read book that I gave you?

- Last night, Anne and I finally went to restaurant that her sister loves.

- Where did time go? It’s already last day of fall break.

- I want to travel across Sahara Desert one day.

- Jake is going to Philippines this summer.

- Have you met Sophie? She’s sweetest kid ever.

- Could you pass me salt?

Part 2. Fill-in-the-blank task (10 items)

Ice cream is ___a____ sweetened frozen food typically eaten as ___a____ snack or dessert. ___The____ meaning of ___the____ phrase "ice cream" varies from one country to another. In some countries, such as ___the____ United States, ___the____ phrase "ice cream" applies only to ___a____ specific variety. Ice cream may be served in ___X____ dishes, for eating with ___a____ spoon, or in ___X____ cones, which are licked.