Dear Reader,

Together with the rest of the editorial team, we are pleased to introduce the fourth issue of the MSU Working Papers in Second Language Studies. The MSU Working Papers in SLS is an open access publication, meaning that content is distributed freely online and there are no restrictions on access.

Before introducing the articles in this issue, we would like to briefly explain the process and purpose of the Working Papers. Ever since the inaugural issue, which was published in 2009, the Working Papers have been organized, written, reviewed, selected, proofread, and edited by volunteers affiliated with the Second Language Studies and TESOL programs at Michigan State University. The purpose is not to publish polished research articles but to provide a forum for students to publish high quality works in progress, book and software reviews, research proposals, and interviews with established researchers.

This issue of the Working Papers includes two interviews, three book reviews, two research proposals, and a research paper.

Each February, the Second Language Studies program holds a research symposium, to which researchers from other universities are invited to give talks. After the 2013 symposium, Betsy Lavolette and Lorena Valmori interviewed Dr. Matthew Poehner, and Ayman Mohammed interviewed Dr. Norbert Schmitt.


The Working Papers also showcases works-in-progress. To this end, we accepted to research proposals this year. Sehoon Jung discusses the effects of task repetition and written corrective feedback in L2 writing. Betsy Lavolette explores the effects of rehearsal on ESL learners’ test tasks.

We also have one full paper. Laura Ballard investigated L1 and L2 English speakers’ ability to identify English accents, and discusses the role that accent familiarity plays in L2 comprehension. She also discusses some of the implications her results have for English language programs employing both native and non-native English speaking teachers.

Finally, in addition to the contributors to this issue, we would like to thank the volunteer copy editors and section editors. Their names are listed below. We are also indebted to the reviewers who provided valuable feedback on drafts of the articles, but who will remain anonymous. We also received invaluable support from Dr. Susan Gass, and the rest of the SLS department, and Russ Werner, who provided tech support for the Working Papers website. Without the help of these tireless volunteers,
the Working Papers would not be possible.

All of the following volunteers are MSU students:

Copy Editors:

• Karen Cheung
• Yaqiong Cui
• Virginia David
• Frances LaMielle
• Ayman Mohamed
• Phuong Nguyen

Section Editors:

• Solène Inceoglu
• Ayman Mohamed
• Le Anne Spino
• Dominik Wolff


Sehoon Jung
Megan Smith
Co-editors
Interview with Matthew Poehner

Interviewed by Betsy Lavolette\(^a\) and Lorena Valmorí\(^b\)
Second Language Studies Program
Michigan State University
\(^a\)betsy@msu.edu, \(^b\)valmoril@msu.edu

Dr. Matthew Poehner is Assistant Professor of World Languages Education and Applied Linguistics at The Pennsylvania State University. He delivered a keynote address at the 2013 Second Language Studies Symposium titled Dynamic Assessment: Understanding Mediation. He was kind enough to speak with us after his talk.

**How did you become an applied linguist and how did you become interested in the field?**

First, as an undergraduate student, I spent some time teaching ESL in the US, and I got really interested in teaching languages. I was studying French at university, and so I thought I could stick around for a while, get some credentials in education, and then become a language teacher. There was no certification at that time in ESL for working in schools, but I did a certification to teach French. I knew I wanted to go back to graduate school, but I wasn’t sure of what I wanted to do. So I applied to a masters program that gave a broad base in literature and civilization and formal linguistics, and I also took a course on SLA. I was really interested in learning more about processes of second language development, and I thought that seemed like it could inform my teaching—because I was still committed to teaching. By the way, this was at Penn State, and they were building a PhD program for people who were interested in applied linguistics. Big names were coming, and I applied to stick around there. I thought that was a good place to do a PhD. One thing led to another!

**Is that how you began working with Dr. James Lantolf?**

Yes, Jim Lantolf came to Penn State, and he offered some courses on sociocultural theory, and I became fascinated by that. Not just as a way of thinking about second language learning, but as a way of thinking about what it means to be a human being because it is a broad ranging theory that speaks to so many different academic areas. It offers a fascinating perspective on being human. So, I was taking some classes with him, and I was taking some classes in language assessment with Dr. Elana Shohamy. I was interested in sociocultural theory, and I was interested in assessment, but these two things seemed like they couldn’t go together. Then, very serendipitously, I was having dinner with Dr. Tim McNamara, and he introduced me to Reuven Feuerstein’s work. I thought that his work sounded exactly like sociocultural theory, and that’s how I got interested in Dynamic Assessment (DA). I realized that people were working in DA who were not
schooled in depth in sociocultural theory. So I tried to speak to the second language people about the Vygotskian view of teaching and assessment, but I also tried to speak to the DA scholars outside of the field in order to show them where the theory originated and what I think Vygotsky had in mind.

**What is the history of DA? Who invented it?**

Vygotsky wasn’t really a developmental or child psychologist. He wanted to develop what he called a “unified scientific psychology.” He was thinking about what the one psychology would be, the parent psychology overarching the various disciplines: a unified theory for what it means to have a human consciousness and human mind, how it develops and is maintained. He was thinking very broadly, but at the same time, he wasn’t only operating at this sort of meta-level. He was also faced with a lot of practical problems. The context in which he was working was the very early days of the Soviet Union, where they were setting up a public school system and trying to integrate children from all walks of life: from urban or country environments, some without any formal schooling, some victims of poverty, and many of them speaking different languages. He was thinking about how to prepare teachers to meet those different needs and about the curriculum. He worked a lot in teacher education and educational psychology. He was dealing with real, concrete problems that needed to be addressed immediately as well as with broad theoretical ideas, and he saw this relationship between theory and practice as the idea that the two are tied up in one another. This is an idea that I’ve been trying to develop in the L2 field with my colleague, Jim Lantolf, rather than separating theory and practice. Take Vygotsky’s argument that a good theory should be able to guide practical activity, but that the practical activity has to be the test of theory. So if the theory doesn’t hold up in practice, then it is not a good theory, and you need to revise it or throw it out. That’s exactly where DA comes from. Vygotsky came up with this idea of how abilities develop as a result of engaging in activities with others and being mediated by others, and talked about it as the Zone of Proximal Development. But this should hold up in practice, so he started working with teachers and having them think concretely about how they could organize activities that could be used to foster and support this development. He did not use the term *DA* himself because his life was cut short tragically, but others who continued on with his work, among them his colleague Alexander Luria, started talking about assessments that are dynamic. The term later was picked up outside Russia in different parts of the world, and the term stuck. The main proponent of DA for years was the Israeli researcher and practitioner was Rubin Feuerstein, who works mostly with children with special needs.

**Is DA problematic for the results of assessment?**

We typically think of assessment as a kind of snapshot, and the image that comes to mind is a test. Testing is the tried and true way, but it is a bizarre scenario. When else in life are you put in a situation where you are given a task, you are not allowed to use resources or ask for help, you have only 45 minutes, it is the same for everybody regardless of background, and we are going to quantify your performance
with a number so that we can compare it with other people’s performance? Only very recently have we been subjected to widespread testing. It is linked to the rise of experimental psychology, where there’s the idea that we can measure mental capabilities in the same way as you measure physical attributes like height, weight, and blood pressure. That’s the script of testing. But DA comes from a really different paradigm. When we start thinking of providing interaction, because we have that testing script in mind, it seems like cheating. Using psychometric terms, the results are going to be contaminated because we want to extract a sample of your knowledge. In DA, we take a very different perspective, which is exactly what Vygotsky was talking about: You can look at independent functioning, as we do in tests, and that can give you a pretty good idea of development that has already taken place. But it is only when you start looking at how people respond and engage in interaction that you can have a sense of abilities that haven’t fully developed but are in the process of emerging. You need both of these things to have a full diagnosis. If you look only at one, you miss a lot of the picture. We can use the example of two kids that when working independently, are both able to solve problems at the level of an eight-year-old. But we noticed that if we start to provide some hints, clues, or feedback during the process, one of these kids is able to improve his performance up to the level of a 10-year-old, and the other up to the level of a 14-year-old. So, are these kids the same, or are they different? If you just look at their independent performance as we do in most testing situations, then they are the same. But if you look at their ZPDs, there are emerging understandings and abilities that are quite different, and these differences are important.

**How can teachers put DA into practice?**

If we look at Feuerstein, he primarily works with children with special needs. They come to his educational center as kids that are educably mentally retarded, and the public school system cannot do anything with them. The expectations for these kids are extremely low. In this kind of situation, standardized testing would reinforce the experience of failing and not being capable. Feuerstein tries to do whatever possible to mediate them, provide support, and to use any kind of instruction in order to see what he can do that will prompt some sort of response in them. That’s his approach to DA. It’s very open-ended and very dialogic. Teachers or mediators will try to do everything possible to try to understand and promote the students’ development.

That is one tradition, and it has been very influential in the second language field. Most of the work that my colleague Jim Lantolf and I have done has been in a classroom setting. Most of the time, teachers are less concerned with affecting test scores. They want to see what can help their students. They want to try whatever can give them better assessment information about their students. They want to pursue interactionist DA, which emphasizes interaction.

The other tradition in DA started in the 1980s in IQ measurement. Cognitive psychologists wanted standardized cognitive measurement, so they wanted standardized tests, and they standardized how they approached the mediation. They would provide multiple attempts for the students to solve the problems. They would give very scripted hints about how
to work through a problem, starting off with something very implicit, like “Try again,” and then “Try again, and this time think about what the problem is asking you to do or how these items are related.” It can become more explicit, like “think about these numbers in the problem. What kind of relationship do they have? ” The idea is that there is a Level 1 hint, a Level 2 hint, and so on. In this way, the mediation can be included in the score. For example, “The student got 80% and needed Level 5 prompts.” We have termed that interventionist DA. It is almost like a treatment in an experiment.

Those are the two traditions. The interventionist tradition is the one we have drawn more from. We asked ourselves, “How can we use DA with second language learners, not just in classroom settings or in tutoring settings? How can we do it in a more formal testing situation? What would it look like in the context of the TOEFL? 15,000 learners are taking the exam, and we cannot give them all open-ended and dialogic DA. We’ve been drawing on the idea of a scripted mediation that is weighted and goes from implicit to more explicit. It is a step-by-step approach in order to standardize the results. Still, the idea is that you’re providing only the support the learner really needs—not too much—because this gives you the best diagnosis.

In DA, you’re not just testing the learners, but you are also teaching them something. If at the beginning of the test the learners learn something, are they likely to do better as they proceed through the test?

Yes, that’s right. In psychometrics, the theory behind the standardized measurement, that is a problem. That is referred to as instrument decay. In order to make a measurement, we have to assume that your abilities aren’t changing in the context of the test. You can’t measure a thing that you can’t pin down. If you get on a scale and your weight is changing, what do you do? You have to assume that it is stable during that process. DA is taking the opposite perspective. If your abilities are changing through the test, that means you’re developing, and that’s the desired outcome. In classroom DA, that is killing two birds with one stone: teaching and assessing simultaneously. One of the ways of tracking this kind of development is looking not just at the independent performance, but looking at how much of support learners need over time. Maybe they started off at the beginning needing a lot of support but later, just a little bit of support. Without DA you’ll miss all of that. When you introduce the idea of transfer, you see if you can apply that principle to a more complex problem. That also gives you a sense of the development that is taking place.

DA seems natural for teachers, but it also seems like it would be natural for placement tests because of the problem of students who are close to the cut scores.

I think that’s right. Using DA, we can place students on the basis of the prediction of learning that’s going to take place. Vygotsky’s theory would suggest that they should be placed on the basis of their ZPD, not just on the basis of their independent performance.

What is the role of students’ motivation in DA?

One of our students was working with first language literacy development in
public schools with kids who were having trouble with the state-mandated standardized tests. One of the things she found using DA over time was that it helped students perform better, but there was an affective motivational change as well. She reported that a child said to her at some point, “Now I understand I have a mind and that I’m able to think about things.” One of the major issues he had dealt with initially was being very impulsive and not having the control to think about what the question was asking. He realized that he could think about things and be successful. These kinds of experiences can actually engender motivation and positive self-perception.

What is the future of DA? What are the challenges?

In my opinion, one area of future work is looking more at some of the applications of DA to formal assessment. What would it look like to administer the TOEFL in a dynamic manner? What would it look like to administer an ACTFL or OPI dynamically? This would be testing the predictions of DA. In this area, we are just scratching the surface with the computerized DA work. The other area is still working in classrooms: Is there a way of organizing instruction that is sensitive to learners’ emerging capabilities while also giving glimpses into what their emerging capabilities are? If there is a framework for doing that, what are the ways we can do it effectively with a variety of languages, a variety of learners, a variety of learners’ levels and school contexts? And that’s where collaboration and partnership with teachers is absolutely crucial. Rather than telling them how to do DA, it needs to be driven by teachers that know the different variables. We’re doing some projects with teachers now to get a better sense of what DA might look like in different classroom contexts. Our early studies were one-to-one DA, and this may not always be realistic in a classroom setting.

Are there any tools available to teachers or organizations who would like to create their own computerized DA?

We have produced a guidebook for language teachers that is available on DVD-ROM, with reflection questions and activities. It is available on the website: http://www.calper.la.psu.edu/dyna_assess.phh. It presents the theoretical principles behind DA as well as models and examples of DA. Teachers should think about how it could be relevant in the context they have.

Thank you very much for talking with us.

Thanks a lot. I was happy to come to MSU.
Interview with Norbert Schmitt

Interviewed by Ayman A. Mohamed
Second Language Studies Program
Michigan State University
mohame44@msu.edu

Professor Norbert Schmitt is a professor of Applied Linguistics at the University of Nottingham in the United Kingdom. He started his career in Japan where he taught English as a Foreign Language and pursued his MA degree. He is mainly known for his work in all aspects of second language vocabulary including pedagogy, acquisition, and assessment. His interest in vocabulary extended through his PhD research at Nottingham. His current research interests include phraseology and formulaic sequences, corpus-based research, and the interface between vocabulary knowledge and the ability to read and listen in English. Dr. Schmitt agreed to talk to us during the SLS Spring 2013 symposium.

**Can you tell us how you got involved in second language studies?**

I was mainly born and raised in a farm in Idaho, and I worked there until I was 26 years old. It all started when I decided to live overseas because simply I liked to travel. It was not planned at all to get into my current career, but I would say maybe because I was in martial arts and I enjoyed helping people learn. Then when I went to Japan and started teaching in a formal manner, I loved it and students liked me as a teacher. By then, I started looking at learning and teaching and thinking about pedagogy and I realized that I had fallen into my career. As I said before, it was really luck because it was not a plan at all.

**What motivated your particular interest in vocabulary learning?**

I know that when I was doing my Masters in Japan, I was interested in many things through my observations of second language learning and pedagogy, but I cannot put it down to a certain teacher or assignment. Perhaps I could not understand grammar or universal grammar or many of the other buzzing theories of interest at that time. I thought that vocabulary would be easy to explain and investigate. However, now that I am into vocabulary it has become obvious to me how complex it can be, even more complex than grammar in many ways. Understanding vocabulary acquisition processes is not as simple as some people would imagine.

**What do you think we do not understand yet about vocabulary acquisition?**

I think the main gap in second language vocabulary in general is that we do not
really understand the acquisition process. Yes, we know that the frequency and amount of exposure matter; we know that some words are more easily learned than others, but what is actually the process, the features and the characteristics? Word knowledge involves not only form and meaning but also collocation, register constraints, grammatical characteristics; all of these in my view represent a continuum. Is there a sequence or a kind of order for this process? In other words, if you know the collocations of a word, does it mean you know a number of senses of it? Does it mean you know different members of the word family? So having a good description of how words are learnt over time is the key answer for this. Do we assume that different words are learnt in different ways? Or is there a general pattern by which some things come before others. If we understood that, it would be much easier to describe pedagogy, and set vocabulary measurement that we can understand and interpret. In general terms, it is the overall description of the vocabulary acquisition process that is still lacking in second language research.

**What are some of the recent research projects that you have conducted or supervised?**

We have a lot of interesting research going on. I have a student from Saudi Arabia who looked at implicit and explicit knowledge of vocabulary. She found it was very difficult to advance implicit knowledge. We all know about explicit knowledge in which you have conscious control of what you are learning. Instructional input and learning output—those are definitely related and you can control them. But how can we enhance the implicit knowledge? She found with priming studies that it was quite difficult with instruction to push forward and advance implicit knowledge of vocabulary. Another student—who is now my colleague and we are glad that we have hired her—has implemented the eye tracking technology in monitoring the vocabulary-learning process, which is unique and exciting. She was actually testing the claims of incidental learning that may take place through extensive reading. To do this, she tracked learners as they interacted with a written text that involved nonwords. One interesting finding that came out of this research was that after about three to four exposures to target words, their recognition becomes quicker, and by the eighth exposure, reading fluency of new words starts to act like the known words. I believe these research projects—among others—are very promising and could be a major step in furthering our understanding of the process of vocabulary learning.

**Coming from a UK research environment, how do you characterize the approach of applied linguistics there as compared to the U.S. context?**

In general, the American approach tends to be very empirical, data driven, theory driven, and experimentally driven. The British approach tends to be more holistic. If you look at grammar for example, you would see them [Americans] talking about UG processes, innateness, and scientific explanation in very specific details. On the other hand, in the UK, grammar is viewed as
meaning-based; you have to understand the context, we have to understand who is doing the communication, what the situation, and what language is used. In testing as well, the British testing is more holistic unlike the U.S. system, which draws heavily on reliability and consistency. Talking about PhD training in both contexts, there is definitely a big difference. In the U.S., students are seen as cheap teaching. In the UK, they are viewed as income generators. Students in the U.S. take longer because they have to teach. They get teaching experience, which is good, but they end up dedicating less time to research. In the UK, students take less time, an average of three years and a half but they end up doing more research than an average student in the American system. I would say the British system is more pertinent to stronger students who can get into research right away without further background. In the US, students get more support through classes and committee supervision. They also get broader background than students in the UK do. I am not saying that one system is better than the other, but it rather depends on the nature of students and their preparation, their specific needs and the best learning environment that matches their abilities.

As a professor and supervisor, what advice do you offer to PhD students in general?

I think one big piece of advice is that if you want to get a job, publications are the key thing that people look for. When you do your research, do not just finish the research and think about writing it later. My key advice is try to write up your research as soon as you are done, in a journal format, and submit it because it gets a lot of time in today’s world, about six months before you get to review and over a year in most cases. If you need some publications in press, submit your work earlier in your PhD and not towards the end of everything. Do interesting research that readers want to read about. This is my other advice. You need to have an identity, something that other institutions would be interested in putting in their program. Think about this in a more practical way.

And what advice do you give for us to be good researchers?

Start with a research question. Do not just think of methodology in a way like I want to use FMRI or eye tracking, without giving enough thought of what you are looking for. Frame your specific question, and then think of the ways, methods and designs that are more likely to answer your questions. This is simply the golden tip for decent research.

From an expert perspective, what are the best journals in vocabulary research?

There is no best journal for vocabulary research. I am happy about that because if there was a vocabulary journal, only vocabulary specialists would read it. However, I want to be an applied linguist, I want my research to be read by language testers, by psycholinguists and other readership, not just by people who are only interested in vocabulary research. Simply, I want my research to be used by other people and I am glad I can publish in a range of journals like SSLA, Language Learning,
Language Teaching research, and many others.

**Before we end this interesting interview, is there anything you would like to add?**

I am happy to be a part of this exciting symposium. It is always a great opportunity to exchange ideas about research and get to know what work is being done in the field. I enjoyed my trip to Michigan State University, and enjoyed meeting people from all over the world. I wish you best of luck in your future endeavors.
The growing number of international students who come to study at American universities and the over 99,000 students who are specifically enrolled in English language programs (U.S. Immigration and Customs Enforcement, 2013) has revealed the need for well-developed and carefully tailored resources that prepare English-as-a-second-language (ESL) students for academic success. This handbook and workbook combination is one of the resources that has been developed to prepare ESL students for success both in the writing classroom and in their academic careers.

According to the preface, the handbook has two intended audiences: ESL students enrolled in academic content classes, and, when accompanied by the workbook, English-for-academic-purposes (EAP) writing students. Several different features of the text clearly show the intention of this book to serve as a resource for ESL students. For example, the handbook covers a range of topics that are relevant to ESL writers. Within these topics, the range is broad and only relevant details are presented, and this does not overwhelm students with unnecessary information. Additionally, in an attempt to make the text more accessible to and relevant for ESL learners, the handbook uses real examples written by ESL students. Likewise, the workbook uses authentic examples that correspond directly to topics covered in the handbook and it provides further exercises that can be used in class or can be assigned as homework.

The handbook is clearly organized into eight sections. Within each section, there are defined sub-sections that give an overview of the information that follows. This overview is also accompanied by detailed descriptions, examples, and exercises. The book also provides an answer key for all of the exercises provided throughout. Table 1 below is an overview of the content of each of the eight sections.
<table>
<thead>
<tr>
<th>Section</th>
<th>Content</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Section 1:</strong> <em>The Writing Process</em></td>
<td>• Writing process overview&lt;br&gt;• Understanding writing assignments&lt;br&gt;• Audience&lt;br&gt;• Idea Generation</td>
<td>Create an awareness of the writing process and build a foundation of expectation for academic writing.</td>
</tr>
<tr>
<td><strong>Section 2:</strong> <em>Essay Structure</em></td>
<td>• Shaping a paragraph and an essay&lt;br&gt;• Revision&lt;br&gt;• 10 steps guide to the writing process</td>
<td>Create an awareness of the components and characteristics of an academic essay.</td>
</tr>
<tr>
<td><strong>Section 3:</strong> <em>Patterns of Essay Organization</em></td>
<td>• Narrative, process, comparison/contrast, cause/effect, and argument essays&lt;br&gt;• Combining patterns of organization</td>
<td>Create an awareness of purpose and organization of common genres.</td>
</tr>
<tr>
<td><strong>Section 4:</strong> <em>Research Paper</em></td>
<td>• Incorporating sources&lt;br&gt;• Drafting&lt;br&gt;• Citation styles</td>
<td>Create an awareness of the accepted conventions and guidelines for academic research papers.</td>
</tr>
<tr>
<td><strong>Section 5:</strong> <em>Grammar and Style</em></td>
<td>• Word level grammar&lt;br&gt;• Sentence level grammar</td>
<td>Provides intervention for grammatical issues.</td>
</tr>
<tr>
<td><strong>Section 6:</strong> <em>Punctuation</em></td>
<td>• All major punctuation</td>
<td>Describes purpose for and provides intervention in usage of punctuation.</td>
</tr>
<tr>
<td><strong>Section 7:</strong> <em>Quick Reference</em></td>
<td>• Connectors&lt;br&gt;• Words often confused&lt;br&gt;• Active/passive voice&lt;br&gt;• Verb tenses&lt;br&gt;• Irregular verb forms&lt;br&gt;• Gerunds and infinitives&lt;br&gt;• Phrasal verbs</td>
<td>Provides accessible charts for intervention in common writing errors.</td>
</tr>
<tr>
<td><strong>Section 8:</strong> <em>Special Types of Writing</em></td>
<td>• Email guidelines&lt;br&gt;• Resumes&lt;br&gt;• Cover letters&lt;br&gt;• Personal statements</td>
<td>Creates an awareness of professionally-oriented writing.</td>
</tr>
</tbody>
</table>
This handbook follows a process approach in which the goal is to inform students about the process of writing and to help them learn to write through selected teacher interventions (Susser, 1994). Sections one, two, and four bring awareness to the process of writing, while sections three and eight bring awareness to conventions and guidelines in academia. Sections five, six, and seven provide resources for intervention in terms of specific writing skills. As part of a teacher-lead (or when using this book as a resource, student-initiated) intervention, this handbook’s grammatical section follows a form-focused instruction (FFI) approach. Within a FFI frame, the book implements a present, practice, produce (PPP) structure that explains a grammar point, provides drills or exercises to practice the point, and then expects its readers to apply the point in the student’s own writing.

Though the five-paragraph essay is highly debated, the handbook uses it as a building block for more complex writing. Rather than serving as a model to copy that leads to “formulaic, empty writing” as Macbeth (2002, p. 35) put it, the five-paragraph essay is meant to model the principles of organization that lead students to successfully present ideas in an essay. This approach to writing can be seen in section two (Essay Structure), which outlines important organizational features through the use of a moldable five-paragraph essay. Other models are given in section three (Patterns of Essay Organization), which breaks down the common features of five major essay genres and gives a full-length example essay for each genre. Even though the handbook provides models, it also analyzes the features and functions and explains the complex and dynamic nature of genre, rather than teaching students to copy a static model.

If it were to be used as the sole book for a course, one of the major drawbacks about this text is the lack of free production activities. The majority of the exercises included in the handbook are highly controlled in nature and do not induce free production. Likewise, the biggest concern with the text is that it does not ask students to produce any substantial writing; it merely guides students in the writing process and provides intervention exercises such as error correction tasks, sentence formulation, and vocabulary use.

As mentioned above, the intended audience for this text is EAP writing students. If it is used with the workbook, the author suggests that the handbook can be used as the sole text for an EAP writing class. Having personally used this text for such a class, I would not recommend using it as the only text for a class, primarily because of the lack of substantial writing production discussed above and from a realization that the text offers a broad range of topics at the expense of in-depth explanations. When preparing lessons, I often feel the need to supplement the text with materials that offer more robust explanations and a greater collection of examples of topics covered in the book. This being said, I do believe that the handbook would more appropriately serve
as a true handbook, and should be primarily used for reference. Because it is navigable and accessible, the combination of handbook and textbook could serve as a valuable resource for college students who are enrolled in advanced intensive English programs, in EAP courses, or in non-ESL writing or content courses at the college level. I also find this handbook useful in one-on-one writing sessions with ESL students.

This text has many strengths, namely that it not a traditional writing textbook, but rather a handbook to be used as a supplement and a resources for students. The workbook increases the usefulness of the book by providing students with additional exercises to implement the skills that are being presented each section. Other features, such as the detailed content section which makes finding information easy, the quick reference which can be referred to while editing a paper, and the many authentic examples, make this handbook a valuable resource for both students and teachers.

References


Review of Academic Encounters Level 4 Student’s Book 
Listening and Speaking with DVD: Human Behavior.

Review by Yangting Wang
M.A. TESOL Program
Michigan State University
wangya36@msu.edu

1. A descriptive summary of the book
The rationale for content-based instruction is that the content is often academic subject matter (Brinton, Snow, & Wesche 2003). Academic Encounters Listening and Speaking: Human Behavior is a content-based textbook dealing with the topic of human behavior related to psychology and communication. The book contains four units: Health, development through life, nonverbal messages and interpersonal relationships. There are two chapters in each unit. The content of each chapter consists most of interviews, lectures and surveys. The interviews and academic lectures are semi-authentic materials from real college classrooms. The DVD that accompanies the text contains all the lectures from the textbook. In each chapter, students are asked to develop four skills: listening, speaking, vocabulary, and note taking. According to the author of the book, the intended population for the book is high intermediate or low advanced students in secondary school. The purpose of the textbook is to provide students with the skills necessary for taking academic courses in English and to prepare them to be successful in their academic study.

Each unit in the book has four parts, and they are distributed across each chapter. These parts are: Getting Started, Real-Life Voices, In your own Voices, and Academic Listening and Note Taking. The first part, “Getting Started,” is a warm up activity that includes a top-down reading activity and a top-down listening activity. Both of these activities are followed by comprehension questions. These activities are consistent with the whole language approach in which “students work from the top-down, attempting first to understand the meaning of the overall text before they work on the linguistic forms comprising it” (Larsen-Freeman & Anderson, 2011, P. 141). The section called “Real-Life Voices” includes one to three authentic interviews. “In your own Voices” gives students a chance to use expressions to ask, answer and perform, share personal information, describe a typical scene, and conduct surveys. “Academic Listening and Note Taking” includes lecture materials and three types of activities. At the end of each unit an oral presentation, which serves to test learning outcomes, is given.
2. An evaluative summary of the book

When it comes to tasks and activities, the author chooses to use a top-down activity first as a pre-reading or pre-listening activity to introduce each topic. The reason for this is to help students build schema, predict the content and anticipate the listening. According to Brown (2011), prior knowledge has a positive effect on comprehension. On the other hand, in during-listening activities, the author changes the focus of the activities to bottom-up tasks such as vocabulary and grammar exercises. Apart from the specific activities and multitasks in the content, the organization of the book is almost the same in each chapter. This gives students a clear view of the content and structure of each chapter and helps them to be more efficient in conducting tasks and learning skills.

At the end of each unit, learning outcomes are evaluated as a whole through word forms, topic reviews and oral presentations. The teacher can thus test students’ speaking and vocabulary skills. Although the evaluation of listening and note-taking skills is not given in the text, the teacher can check students’ mastery of these skills by assigning homework. For instance, the teacher can require students to watch the DVD and complete listening exercises, or require students to hand in their listening notes to evaluate these two skills.

The book is successful in three aspects. It is comprehensive in that it contains not only speaking and listening tasks, but also vocabulary and note-taking tasks; it is abundant as it provides not only semi-authentic materials, but also knowledge of academic content; it is interesting in that colorful symbols, graphics, and humorous pictures are included to support the discussion. However, I have one concern about its effectiveness in improving students’ overall speaking and listening skills. This book seems to assume that students can improve their speaking and listening skills by conducting surveys and working in pairs to fill in tables and charts. This means that in order to complete the exercises, students only need to write several words in the missing blanks. Completing these tasks may be interesting, but with vocabulary knowledge, students cannot be pushed to be very communicative and such tasks may require a large amount of time.

3. Reference to current theory and research

There is ample current theory regarding the significance of content, skills and technique addressed in this book. This book reflects several aspects of task-based language teaching. In task-based language teaching, students are asked to do project work (oral presentations), information gap activities (filling in blanks, recording numbers), opinion-gap activities (sharing own personal perspective), and reasoning-gap tasks (inferring information) (Larsen-Freeman & Anderson, 2011). Secondly, there is a growing emphasis on both bottom-up and top-down skills in current perspectives of teaching listening and speaking (Hinkel, 2006). This book successfully integrates these two skills into every task. Thirdly, graphic organizers, which are a technique that requires students to write down missing information on charts, tables and graphics, are also used in the
book. “The Graphic Organizer helps students to organize and remember new things. They combine words and phrases, symbols, and arrows to map knowledge. They include diagrams, tables, columns, and webs” (Larsen-Freeman & Anderson, 2011, P. 142).

Above all, the author integrates current theories of task-based language teaching and content-based instruction into the book which makes it very well suited for low-advanced learners. However, because of the limited time in class, it is always helpful for teachers to choose to present the main parts, to create a cohesive and clear syllabus, and to assign communicative and other remaining tasks to students after class. In this way, students can enhance their overall skills and teachers can better evaluate their learning outcomes.

References


The goal of this book is to combine data-driven findings in the area of Arabic language acquisition with a focus on the order of acquisition, developmental stages and learning patterns of linguistic structures of Arabic. The author starts with research questions on the variables involved in learning Arabic including the developmental path, the role of transfer, the effect of instruction, and other factors involved. Methodologically, the book draws on a pool of longitudinally collected oral data from 9 learners of Arabic and a cross-sectional sample of 109 learners with different native languages including English, French, Spanish, and Japanese. Data were elicited through information gap activities and picture description tasks. Based on analyses from this data, chapters are systematically developed to address topics in separate chapters for discussing the acquisition of specific target features.

In the first chapter, the author provides a brief account of the target structures relevant to the discussion points in later chapters. He gives a concise explanation of word structure, phrasal agreement, verbal structure agreement, gender agreement, and negation patterns. The author selected these structures in particular based on a common observation that although these features are taught quite early in language classrooms, some learners may not acquire them in the expected sequence.

The second chapter counts as a literature review of the existing research in Arabic SLA. The review focuses on studies that investigated the order of acquisition of certain Arabic structures under the Processability Theory framework (Pienemann, 1998; 2005) and through the UG perspective under the Full Transfer/Full Access proposal (Schwartz & Sprouse, 1996). However, it seemed from the review that the studies that actually existed were very few in number and very limited in scope. This was one main reason that justified the author’s stance and highlighted the significance of the book.

The empirical portion of the book, which is presented from chapter three to chapter six, relies on both longitudinal and
cross-sectional data of one wide-scale study. In each chapter, the author clearly lays out the procedures, tasks and analyses in a way that is easy to understand especially for readers who do not have an Arabic background. Chapter three presents the findings on nominal agreement, verbal agreement, and demonstrative agreement. Chapter four discusses the acquisition of tense/aspect and verbal agreement based on a cross-sectional data from English L1, Spanish L1, and Japanese L1. Chapter five focuses on the acquisition of null subjects using the same cross-sectional data while chapter six is devoted to the features of negation, case and mood. The author also provides a content analysis of the textbooks used in Arabic program highlighting the order of presentation for the target structures.

With chapter seven, the author concludes his empirical finding by presenting theoretical underpinnings that could account for the trend of results shown in previous chapters. The bulk of discussion included L2 learnability and processability, L1 transfer and UG access, ultimate attainment and near nativeness, the split INFL hypothesis, and more theoretical implications on UG access and transfer. Although some results provided support for Processability Theory, the author assumes that the Full Transfer/Full Access model is the best representative of the pattern of results found with only minor exceptions that may fall under the Speech Processing model.

In the final chapter, the author provides realistic and useful implications related to material preparation, instruction, teacher preparation, and proficiency testing. By acknowledging research findings in this area, the field of Arabic applied linguistics and pedagogy could be pushed further to find its place among SLA mainstream publications. The author concludes by a call for more research on different populations and in different learning contexts.

In general evaluation, the book is valuable in terms of its topic on Arabic SLA, which has been less well represented in the literature because it is a less commonly taught language. Given the current increasing demand to learn Arabic, especially in the US, a parallel concern was raised about ways of accommodating this need and developing strategies, material, and methods to best serve this purpose. However, the field is still suffering from lack of connection between materials writers, Arabic teachers, professionals, researchers, and testers. This created a gap both in research and pedagogical practices. This book comes out in response to this need for the integration of research and pedagogy.

One advantage of the book is that it combines the bulk of research related to the topic, providing a starting point for further research. Although few studies existed, the author manages to give the reader a sufficient background about the topic. The author organizes research foci by chapters, which makes it easier for the readers to follow and even to pick the topics that are more interesting to pursue. The chapters are not dependent on each other, so that a reader can simply read a chapter as a separate research article. The discussion of
Theoretical models seem to be simple and appropriate to the novice reader who does not have to refer back to outside sources for additional theoretical background. Methodologically, the author/researcher relies on a broad pool of cross-sectional and longitudinal data that he distributes around chapters and presents relevant results according to topic. The variety of the sample is noticed as the data included participants from US, Spain and Japan. Technically, the experimental procedures are presented in a manner that allows replication as the author included screenshots of the material used for the oral elicitation tasks. Results are presented clearly through a very good integration of tables and graphs.

On the other hand, the book is limited in scope in the sense that it is mainly based on the researcher/author’s own work with very few references to other studies. This is justified, however, by the fact that there is really a lack of research in this particular area and that the author’s purpose is to encourage future research by supplying detailed and categorized portions of his own work in an appealing style. At the end of the book, it is evident that it is difficult to arrive at conclusive answers to the research questions posed in the first chapter, but the author did his best in approaching the problem and providing sound explanations from several theoretical backgrounds.

The greatest contribution of this book is to make this research available with the goal of integrating it into mainstream SLA studies, challenging the current state of alienation that Arabic applied linguistics is facing. It is a comprehensive resource for teachers, textbook authors and testers, and most importantly researchers with Arabic linguistics background who can bridge the gap between theory and practice and respond to the current demands of the field.

References


The Effect of Task Repetition and Corrective Feedback in L2 Writing: A Pilot Study

Sehoon Jung
Second Language Studies Program
Michigan State University
jungse@msu.edu

There is great interest in the role of task planning on learner performance and L2 development (for a review, see Ellis, 2009). In this line of research, the question of whether or not repetition of the same oral/written task—repeating as a type of task planning—is pedagogically useful for second language (L2) learning is an intriguing one both for second language researchers and educators. In particular, it is of empirical and practical interest to explore to what extent task repetition helps learners to produce improved language output with less cognitive burden inside and outside of the repetition cycle, and how this contributes to the development of multiple dimensions of L2 spoken or written proficiency in the short- and long-term. With this in mind, findings from previous research suggest that implementation of task repetition with or without additional treatment (e.g., reformulation, or feedback) may be beneficial for learners to some extent (e.g., Adams, 2003; Bygate, 1996, 2001; Bygate & Samuda, 2005; Gass, Mackey, Alvarez-Torres, & Fernandez-Garcia, 1999; Lynch & Maclean, 2000; Sheppard, 2006; Swain & Lapkin, 2002).

However, such positive effects reported in the literature have often come with limitations regarding their range of benefits. First, the learners generally showed improvement on fluency and complexity, but showed less or no improvement on linguistic accuracy (e.g., Larsen-Freeman, 2006). The second limitation of the effect is that the positive effects found during the repetition cycle were less likely to be extended to learners’ subsequent performance in a new task. Taken together, these limitations suggest that no clear evidence is yet available that task repetition can ultimately lead to L2 development and/or acquisition of certain linguistic forms. There have been a few suggestions proposed to clarify the unclear relationship between task-repetition and L2 acquisition. For example, Bygate (2001) recommended that a large amount of practice through repetition may be necessary in order for repetition to have a significant effect on accuracy. Ellis (2009) provided an alternative suggestion, which is that it may be necessary for learners to receive some type of feedback on their initial performance before repeating it.
Taking these limitations and suggestions into account, the current study aims to investigate the effect of task repetition, corrective written feedback, and the interaction between the two on the L2 academic writing performance by L1 Korean ESL learners, specifically on their accuracy, complexity, and fluency. It is also important to note that, to date, task repetition research has heavily focused on learners’ performance on oral tasks with little attention to performance on written tasks (Robinson, 2011). Given this lack of research on the effect of task-repetition on the writing performance, it is another goal of this study to fill this gap in the literature.

**Theoretical background**

The theoretical background of this study is closely associated with the following questions:

- What specific processing procedures do writers go through to write?
- What are the potential cognitive obstacles L2 learners might encounter as they write and what are the consequences of these obstacles on their complexity, accuracy, and fluency?
- How can task-planning, in particular, task-repetition, assist learners to overcome such obstacles?
- How can the provision of written corrective feedback help learners to improve their writing performance?

According to Kellogg’s (1996, 2001) cognitive model of writing process, the writing process entails 3 components consisting of 6 basic sub-processes: formulation [planning & translating], execution [programming & executing], and monitoring [reading & translating (see Table 1 for a description of the components). Kellogg also discussed the role of working memory in this model. That is, while the three components are activated concurrently during the process of writing, all the sub-processes (except the executing3 sub-process) need to compete for a working memory that includes limited attentional resources in the central executive (the control system of limited attentional resources) in Baddeley’s (2003) working memory system. Therefore, it may be necessary for L2 writers to manage their limited attentional resources effectively in various competition contexts during ongoing interactions of the sub-processes4.

As a number of SLA researchers have suggested, however, it seems very likely that the competitions between the process components may impose a heavy processing burden that exceeds learners’ limited attentional capacities, especially with time pressure (e.g., Ellis, 2005; Skehan, 1996; Foster & Skehan, 1996). In the same vein, Skehan (1996) further argued that three aspects of language production—namely accuracy, complexity, and fluency—may be likely to be forced to compete with one another to compensate for limited processing ability, possibly resulting in a trade-off effect. (e.g., better accuracy at the expense of complexity, or fluency; but see also Robinson, 2001, who proposed that complex nature of a task would increase both accuracy and complexity).
Table 1. Kellogg’s (1996) model of writing process

<table>
<thead>
<tr>
<th>Process</th>
<th>Steps</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formulation</td>
<td>Planning: Goal setting, idea generation, and organization</td>
</tr>
<tr>
<td></td>
<td>Translating: Selecting linguistic items at the lexical &amp; syntactic level</td>
</tr>
<tr>
<td>Execution</td>
<td>Programming: Conversion of translated item into the output modality</td>
</tr>
<tr>
<td></td>
<td>Executing: Production of sentences</td>
</tr>
<tr>
<td>Monitoring</td>
<td>Reading: Reading of his/her written texts for review</td>
</tr>
<tr>
<td></td>
<td>Editing: Editing at the micro- and macro-level</td>
</tr>
</tbody>
</table>

While the limited attentional capacity is obviously one of the potential obstacles language learners face in the process of their writing, a number of studies have examined whether the provision of a certain type of task planning can compensate for learners’ limited processing abilities and thus assist them to produce improved L2 output (Ellis, 2005). This study focuses on the effect of task-repetition, among other types of task planning.

**The role of task repetition**

Task-repetition studies (both oral and written) thus far have primarily focused on two issues: whether learners show better writing performance as they perform the same writing task a second time, and whether learners show better writing performance in a new task after they participate in a task-repetition activity. One study that tested these two issues together is Gass et al. (1999), which assumed that increased familiarity of the content from the repeated oral task would provide learners with a better opportunity to attend more to linguistic resources. In their study, English speaking learners of Spanish watched video clips 3 times with 2 to 3 days' break in between, and then watched a new video clip after 1 week. In simultaneous oral production tasks while watching the video clips, learners in the task-repetition group showed improvement in overall proficiency, as it was measured via the magnitude estimation procedure, morpho-syntax (limited to the Spanish copulas _ser_ and _estar_), and lexical complexity; however, such improvement did not carry over to a new task when the learners were given a new video clip for oral production. Gass et al. speculated that learners in the same content group (i.e., repetition group) in their study might have lost their interest at some point during repetition (but, see Hawkes, 2009, for the opposite claim regarding the relationship between task-repetition and motivation). Bygate (2001) also reported similar results in terms of the task-repetition effect on complexity, fluency, and accuracy; improvement in fluency and complexity, but not accuracy when it was measured by average number of errors per t-unit.

There are also a few studies that used both written and oral production modalities in the rehearsal and repetition respectively, or vice versa. Larsen-Freeman (2006), for example, explored 5 L1 Chinese-speaking ESL learners’ performance on a writing task, then oral narrative repetition tasks over a 6-month period. While the author found performance variability between individual
learners across time, the results from the quantitative analyses revealed that accuracy (operationalized as error-free t-unit/total t-unit ratio) decreased when the participants repeated the task for the second time. In contrast to this finding, Ellis (1987), who also implemented a mixed written-oral repetition task, found a positive effect of task repetition on accuracy, specifically on English regular past tense. However, this study differs from Bygate (2001) and Larsen-Freeman (2006) in that the oral repetition task was given immediately after the written task of the same content. As a result, it was not clear whether the improved accuracy was due to the virtue of short-term memory or by acquisition of the form.

Taking these issues into consideration, this study aims to investigate the effect of task repetition, corrective feedback, and interaction between the two on ESL learners’ academic writing performance. The research questions of the study are as follows:

**Research Questions**

1. Does fluency, complexity and accuracy increase when learners repeat the same L2 academic writing task?
2. Does corrective feedback on learners’ academic writing result in increases in fluency, complexity and accuracy in a repeated task as well as in a new writing task?
3. Does repeating the same writing task and receiving corrective feedback result in increased fluency, complexity and accuracy in a new writing task?

**Method**

**Participants**

Participants in this pilot study were 8 L1 Korean-speaking learners of English as a second language (ESL) who were enrolled either in an ESL program (4 participants) or a degree program (1 undergraduate and 1 graduate) at a large Midwest university in the United States (see Table 2). Two of the participants attended only the first session and did not return for the later sessions, and consequently only the results obtained from the 6 participants are reported in this paper. English proficiency of the participants was evaluated in two ways: Participants’ self-rated proficiency in writing and the administration of the Grammar section of MTELP (Michigan Test of English Language Proficiency). The reported self-rated proficiency of the participants ranged from low-intermediate to high-intermediate, and the MTELP results showed variation to some degree across the participants (50-72.5% accuracy on 40 test items). Participants were also asked to take an online typing test (www.typingtest.com) for two minutes to ensure their typing skills are comparable to each other. After completing the proficiency test and the typing test, the participants were then randomly assigned into 4 groups; Group 1 (Repetition with feedback), Group 2 (Repetition with no feedback), Group 3 (No repetition with feedback), and lastly Group 4 (No repetition with no feedback).
Table 2. Summary of the participants’ bio-data, proficiency, and typing speed results

<table>
<thead>
<tr>
<th>ID</th>
<th>Condition</th>
<th>Age</th>
<th>Self-rated Proficiency [Writing]</th>
<th>Grammar Score (%)</th>
<th>Typing Speed (wpm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>101</td>
<td>R-FB</td>
<td>21</td>
<td>Intermediate</td>
<td>24 (60.0)</td>
<td>36</td>
</tr>
<tr>
<td>102</td>
<td>R-FB</td>
<td>24</td>
<td>Low-intermediate</td>
<td>20 (50.0)</td>
<td>40</td>
</tr>
<tr>
<td>103</td>
<td>R-NFB</td>
<td>33</td>
<td>High-intermediate</td>
<td>29 (72.5)</td>
<td>45</td>
</tr>
<tr>
<td>104</td>
<td>R-NFB</td>
<td>22</td>
<td>Low-intermediate</td>
<td>19 (47.5)</td>
<td>39</td>
</tr>
<tr>
<td>105</td>
<td>NR-FB</td>
<td>20</td>
<td>Intermediate</td>
<td>28 (70)</td>
<td>43</td>
</tr>
<tr>
<td>108</td>
<td>NR-NFB</td>
<td>25</td>
<td>High-intermediate</td>
<td>26 (62.5)</td>
<td>47</td>
</tr>
</tbody>
</table>

Notes. R: Repetition / NR: No Repetition FB: Feedback / NCFB: No Feedback

Task & Procedure

All research activities took place on campus in a research lab equipped with a PC where individual participants composed an essay using the designated computer. Three writing topics were selected from a list of iBT TOEFL (Test of English as a Foreign Language) independent essay writing prompts provided by Educational Testing Service (ETS). Topics A and B were both topics in which participants had to discuss whether they preferred to get up early or late (Topic A), or whether they preferred living in a big city versus a small town (Topic B). Topic C was an argumentative prompt, and participants were asked to discuss the role of television in human communication. (See Appendix A for the complete prompts). The writing task was administered using Microsoft Word word-processing software and the participants were instructed to complete their essays in 30 minutes. Participants’ writing process was recorded simultaneously by the screen recording software Camtasia Studio (Ver. 8.0, TechSmith Corp.) for qualitative analyses. The overall procedure is summarized in Figure 1.

On day 1, the two repetition groups (Group 1 & 2) composed an essay about Topic B, which they wrote about again in Session 2 on Day 7 (i.e., task repetition). The other two groups wrote an essay on Topic A on Day 1, and then they completed an essay on Topic B on on Day 3 (i.e., no task repetition). The third composition on Topic C could not be administered in this pilot study mainly due to time constraints.

A review and revision session either with or without feedback was administered on Day 3 between the two writing sessions. The type of feedback in this study was unfocused direct corrective written feedback with brief meta-linguistic explanation on their lexical, morphological, and syntactic errors. No feedback was given to the participants regarding its logicality, idea, or organization. The two feedback groups (Group 1 and Group 3) were asked to review the given feedback and then revise their essays for 15 minutes without access to the feedback they reviewed. In the meantime, the no-feedback groups (Groups 2 & 4) were asked to review and revise their own written essays (with no feedback) for the same amount of time.
Dependent variables

Following the convention of task-planning research, three dimensions of L2 proficiency—accuracy, (grammatical) complexity, and fluency—were measured respectively as dependent variables for the main analyses. First of all, accuracy in this study was operationalized as the ratio of the number of error-free clauses to the total number of clauses in the essay (e.g., Foster, 1996; Foster & Skehan, 1996). Additionally, the average number of errors per 100 words was separately calculated as a supplementary measure of accuracy (e.g., Mehnert, 1998; Sanguran, 2005). The rationale for this addition is that the error-free clause ratio does not account for frequency of errors in a clause. Secondly, grammatical complexity was operationalized as the average number of clauses per t-unit (e.g., Larsen-Freeman, 2006). All accuracy and grammatical complexity coding procedures, including the identification of ‘clause’ and ‘t-unit’, were guided by Polio’s (1997) “guideline for t-units, clauses, word counts, and errors” (p. 138-140). Lastly, the measure of fluency in the present study was operationalized as the average number of words produced per minute (e.g., Mochizuki & Ortega, 2008). Along with these three dependent measures, two raters rated participants’ essays in a holistic way using the iBT integrated writing rubrics (scoring standards), which was to examine how the two independent variables (i.e., repetition and feedback) affect overall quality of the learner essays.

Data Analysis (not reported in this paper)
Prior to the main analyses, it may be necessary to check how comparable the English writing proficiency levels of the 4 groups are at the onset of their participation. For this reason, 4 separate sets of one-way ANOVAs with each dependent variable (i.e., accuracy, complexity, and accuracy) will be conducted in order to determine if proficiency should be treated as covariate in the main analyses.

For the main analyses, two sets of 2 x 2 x 2 mixed design ANOVA will be carried out on each dependent variable: a 2 (task repetition condition) x 2 (corrective written feedback condition) x 2 (time: Day 1 and Day 7) mixed design ANOVA to answer RQ 1 and RQ 2, and a 2 (task repetition condition) x 2 (corrective written feedback condition) x 2 (time: Day 7 and Day 14) mixed design ANOVA answer RQ3.

Results
I was not able to conduct any statistical analyses to answer the research questions, mainly due to the small sample size ($n=6$) and the fact that the experimental procedures were incomplete in that they included only Session 1 and Session 2 as a part of the pilot study, and not Session 3. For this reason, only descriptive analyses of individual participant performance are reported in this paper, specifically their accuracy, complexity, and fluency on the first two essays they composed in Session 1 and Session 2.

Accuracy
The two raters (the researcher and one native English speaker who has years of ESL teaching experience) identified any lexical (e.g., missing article, or misuse of prepositions), morphological (e.g., missing plural or tense marker), and syntactic errors (e.g., number agreement, inappropriate word order, or sentence fragments) on each essay. After the error coding, a total of 399 clauses were identified from the whole essay data. The coding results showed 88.97 percent agreement on error-free clause identification, which included if the raters were in agreement about whether or not a) a clause was error-free, and b) the raters identified the same number of errors from each clause.

Table 3 provides a summary of written performance in terms of learners’ accuracy. Overall, it seemed that the accuracy data showed no clear effect of the two independent variables of the study, namely task-repetition and corrective written feedback. Although three out of the four participants showed slightly improved accuracy in Session 2, two participants in the no-repetition groups (#105 and #108) displayed even greater improvement (18% and 13% increase in accuracy ratio, respectively) through their writing in Session 2. Similarly, the effect of feedback also seemed to be limited since all groups showed improvement regardless of the feedback condition. The accuracy patterns were similar when it was assessed using the average number of errors per 100 words.
Table 3. Summary of accuracy results of learners in Session 1 and Session 2

<table>
<thead>
<tr>
<th>Subject ID</th>
<th>Group</th>
<th>Error-free C / Total C</th>
<th>Accuracy ratio</th>
<th>No. of errors per 100 words</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Session 1</td>
<td>Session 2</td>
<td>Session 1</td>
</tr>
<tr>
<td>101</td>
<td>R-FB</td>
<td>16 / 33</td>
<td>14 / 35</td>
<td>0.48</td>
</tr>
<tr>
<td>102</td>
<td>R-FB</td>
<td>16 / 24</td>
<td>15 / 21</td>
<td>0.67</td>
</tr>
<tr>
<td>103</td>
<td>R-NFB</td>
<td>23 / 44</td>
<td>31 / 53</td>
<td>0.52</td>
</tr>
<tr>
<td>104</td>
<td>R-NFB</td>
<td>9 / 39</td>
<td>16 / 41</td>
<td>0.23</td>
</tr>
<tr>
<td>105</td>
<td>NR-FB</td>
<td>11 / 27</td>
<td>17 / 29</td>
<td>0.41</td>
</tr>
<tr>
<td>108</td>
<td>NR-NFB</td>
<td>10 / 27</td>
<td>13 / 26</td>
<td>0.37</td>
</tr>
</tbody>
</table>

Note. R: Repetition / NR: No Repetition FB: Feedback / NCFB: No Feedback

Complexity

The average number of clauses per T-Unit was calculated to observe how different task repetition and corrective feedback conditions affect learners’ written production in terms of complexity. A clause in this study was defined as a clause that contains both the overt subject and finite verb, following Polio (1997). Consequently, any clause including ellipses, tag questions, and infinitives were not counted as a clause. Other clause types included are as follows: noun clauses (e.g., it is easy to for people to think that a big city provides diverse places...: Subject #104, Session 2), relative clauses (e.g., they may have more mistakes than others who works constructively: Subject #108, Session 1], and adverbial dependent clauses (e.g., When I study in a quiet place, I can concentrate more on my reading: Subject #102, Session 2).

The comparison of complexity across the participants showed somewhat interesting results, in that the participants who wrote an essay on the same topic in Session 2 (i.e., repetition) showed slight increases in their complexity index with the exception of Subject #102, whereas the participants who wrote an essay on a different topic in Session 2 showed slight decreases (see Table 4).

Table 4. Summary of grammatical complexity results of learners in Session 1 and Session 2

<table>
<thead>
<tr>
<th>Subject ID</th>
<th>Group</th>
<th>Average number of clause per T-Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Session 1</td>
</tr>
<tr>
<td>101</td>
<td>R-FB</td>
<td>1.59 (.89)</td>
</tr>
<tr>
<td>102</td>
<td>R-FB</td>
<td>1.33 (.59)</td>
</tr>
<tr>
<td>103</td>
<td>R-NFB</td>
<td>1.15 (.37)</td>
</tr>
<tr>
<td>104</td>
<td>R-NFB</td>
<td>1.26 (.57)</td>
</tr>
<tr>
<td>105</td>
<td>NR-FB</td>
<td>1.38 (.61)</td>
</tr>
<tr>
<td>108</td>
<td>NR-NFB</td>
<td>1.18 (.39)</td>
</tr>
</tbody>
</table>
Fluency

As stated earlier, fluency was operationalized as the average number of words that were produced per minute. One potential confound when using this measure is the length of words. One way of resolving this potential issue is using the average number of syllables per minute (e.g., Ellis & Yuan, 2004; Yuan & Ellis, 2003). Alternatively, however, I calculated the average length of words by dividing the number of characters (including spaces between words) in each essay by the total number of words. The results from this procedure showed no clear difference among the participants in terms of the mean length of words they produced; the mean length of individual words across the participants ranged from 4.3 to 4.6 characters per word in Session 1, and from 4.2 to 5.1 characters per word in Session 2. Based on this analysis, word length did not seem to function as a compounding variable when fluency was measured using the number of words.

The comparison of fluency data among the participants showed some positive effect of task repetition. Indeed, those who repeated the same task (subject #101, 102, 103, and 104) showed relatively greater improvement than those who did the different task (subject #105, 108) (see Table 5). However, no clear effect of feedback was found, in that all participants displayed similar ranges of improvement in Session 2.

<table>
<thead>
<tr>
<th>Subject ID (Group)</th>
<th>Total number of words</th>
<th>Composition Duration (minutes)</th>
<th>Number of words per minute</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Session 1</td>
<td>Session 2</td>
<td>Session 1</td>
</tr>
<tr>
<td>101 (RFB)</td>
<td>365</td>
<td>383</td>
<td>27</td>
</tr>
<tr>
<td>102 (RFB)</td>
<td>253</td>
<td>270</td>
<td>30</td>
</tr>
<tr>
<td>103 (RNFB)</td>
<td>411</td>
<td>531</td>
<td>30</td>
</tr>
<tr>
<td>104 (RNFB)</td>
<td>271</td>
<td>311</td>
<td>30</td>
</tr>
<tr>
<td>105 (NRFB)</td>
<td>376</td>
<td>377</td>
<td>30</td>
</tr>
<tr>
<td>108 (NRNFB)</td>
<td>250</td>
<td>276</td>
<td>30</td>
</tr>
</tbody>
</table>

Discussion

The first goal of this pilot study was to investigate whether ESL learners can produce improved written output and minimize trade-off effects, if any, if they are given a chance to rehearse the written task (Session 1) before performing the main task (i.e., repetition in Session 2). The preliminary observations of the descriptive data showed at least some positive effect of task repetition. For example, the two participants in the no-repetition groups (subject #105 & #108) showed some trade-off effects between Session 1 and Session 2. That is, the no-repetition group displayed
slightly lower complexity indexes in Session 2 compared to those in Session 1, whereas their performance regarding accuracy and fluency improved slightly.

On the other hand, the participants who repeated the same task in Session 2 produced slightly more complex sentence structures more accurately and more fluently. One exception was subject #102, who showed improved accuracy (from 67 to 71 percent of error-free clauses), but displayed lower complexity (from 1.33 to 1.17 clauses per minute) at the same time. One possible explanation for this exception would be that it is not clear if this participant repeated the same task twice. Note that the writing task in this study provided the participants a chance to choose one of the two options from the prompts (e.g., Option 1: prefer to live in a small town versus Option 2: prefer to live in a big town). With this in mind, the other three participants in the repetition condition picked the same preference options both in Session 1 and Session 2, whereas subject #102 selected different preference options between Session 1 and Session 2. Therefore, although the subject might have been more familiar with the topic itself, the contents of the essay in Session 1 were very different from those in Session 2.

This study also aimed to explore the effect of corrective written feedback. The preliminary results from the 6 participants, however, failed to provide clear evidence of a feedback effect. To be more specific, although those who received feedback showed slight improvement in accuracy, those who did not receive any feedback also showed a comparable amount of improvement in Session 2 at the same time within the same repetition conditions (i.e., a comparison of subjects #101-102 and #103-104, and a comparison of subjects #105 and #108). One possible reason for such results would be that the participants might not have received sufficient time to review the given feedback. During the review session, the participants in both the feedback and no-feedback conditions were given only 15 minutes to review their essay with/without feedback and then revise them. Furthermore, the type of corrective feedback implemented in this study was unfocused direct corrective feedback plus meta-linguistic information. Given this lack of time and the type of feedback, the participants might not have had enough time to pay their attention to a number of unfocused target forms before they started revision. Therefore, in future research, it may be better to select specific target structures such as English articles or tense/number agreement and provide more focused feedback on the targets.

Based on the results reported in this pilot study, there are several points that need to be reconsidered and changed to further develop the research design and procedure. First of all, as mentioned earlier, one participant who was in the repetition group did write his second essay choosing a different preference option. Therefore, it may be helpful to provide more elaborated instructions asking them to choose the same option as they write the essay for the second time.

Secondly, it may be more accurate and informative to operationalize and measure the three dependent variables—complexity, accuracy, and fluency—in multiple ways.
For example, the accuracy result of subject #101 showed a decrease from 48 percent to 46 percent when it was operationalized as the error-free clause ratio; however, when accuracy was defined as the number of errors per 100 words, the result demonstrated the opposite pattern in that the subject showed slight improvement (from 6.3 to 5.8 errors per 100 words). Therefore, measuring the variables in more than one way would provide more precise results.

Lastly, I would assign a shorter composition time in the future research. In this study, I gave the participants a 30 minute time limit to complete their essay, following the ETS administration guidelines. However, I found at least two potential problems of this way of administration. First, considering the fact that some participants have to write the same essay more than once, the tasks may be too boring for them. In fact, two out of the four participants who repeated the task (subjects #101 and #102) spent only 23 and 28 minutes respectively on the task in the second session whereas the others spent the full amount of time. Furthermore, subject #102 reported that he wrote his second essay choosing a different preference option because he felt it was boring to write the same content again. Another potential problem is that both Kellogg’s model of writing and Skehan’s cognitive model assume limited attentional capacities in the writer’s working memory system under time pressure. Therefore, it seems to be difficult to observe natural aspect of cognitive function with less time pressure since participants’ limited attentional resources will be likely to be compensated by the virtue of more time.

Shortening the composition time also will make it possible for the researcher to add more repetition cycles in the research design, which I believe would more clearly show the effect of task repetition (if any; e.g., Lynch & Maclean, 2000; Sheppard, 2006, as cited in Ellis, 2009), as well as the effect of corrective written feedback.

Lastly, I hope to add a part of qualitative analyses in the future research. In particular, qualitative analyses on how participants responded to the target structures from the feedback during their revision will be able to provide crucial evidence regarding the effect of corrective feedback. In addition, observing recorded writing process will also provide meaningful information. For example, examining whether or not learners make longer pauses at a point where they have to use certain structures that were corrected previously would be able to inform if they pay more attention to the form (regardless its accurate usage). It will also be useful to examine if learners produce certain chunks of words more fluently (e.g., shorter pauses) when they repeat those chunks during the repetition session(s).

**Conclusion**

This study aimed to investigate whether task repetition in second language writing can help learners to manage their limited attentional resources more effectively, and thus produce improved written output in a repeated task as well as in a new task. The preliminary analyses of the descriptive data suggested that the L2 learners performed better at least within the task-repetition cycle and they showed no trade-off effect as their
performance was assessed by accuracy, complexity, and fluency. The question of whether such positive effects can be transferred to a new task is left for future research.

However, the current study failed to find any effect of corrective written feedback, which might have been partly due to the limited time assigned to the participants for the review of the written feedback. In addition, under such time constraints, unfocused feedback on a number of unspecified structures might have made it difficult for the participants to manage their attention to the target structures effectively. Finally, keeping all the issues and problems discussed in this paper in mind, I hope to further develop its research design and procedures.

1 By (performance) outside of task-repetition, I mean performance on a new task followed by task-repetition activities.
2 Note that discussion of such a pattern on proficiency measures is limited to the task-repetition type of task planning. Foster (1996) found increased accuracy when participants were given unguided task planning time. Yuan and Ellis (2003) and Ellis and Yuan (2004) also found increased accuracy when the planning type was unpressured online-planning.
3 Note that Kellogg’s writing model assumes writing in L1. For native speakers, it might not be necessary to bring attention for the execution of sentence production, presumably because their handwriting or typing is fully automatized in general. Ellis (2005), however, argued that for some L2 learners, especially for those whose L1 uses different scripts from L2, may need to pay extra attention in this stage.
4 See also Hayes’s (2006) writing model that added social/environmental factors into the model, in addition to working memory and long term memory.
5 Note, however, that those two targets, according to the authors, had been a heavy focus of instruction in class prior to students’ participation in the study. Thus, it is not clear if the improved accuracy on this specific grammar component is due to task repetition or significant instruction.
6 To my best knowledge, no previous studies in L2 writing measured the typing speed when the modality of writing was by computer. However, different levels of typing skills may potentially function as an intervening variable, particularly with respect to fluency. In addition, relatively poorer/slower typing skills will affect participants’ performance on accuracy and complexity as well because they have to take up more attentional resources for typing, compared to those whose typing skill is better (Ellis, 2005).
References


writing (pp. 57-71). Mahwah, NJ: Lawrence Erlbaum Associates.
Appendix A. ETS iBT Writing Topics

[Topic1. Preference 1 — Life style]
Some people prefer to get up early in the morning and start the day’s work. Others prefer to get up later in the day and work until late at night. Which do you prefer? Use specific reasons and examples to support your choice.

[Topic2. Preference 2 — Living style]
Some people prefer to live in a small town. Others prefer to live in a big city. Which place would you prefer to live in? Use specific reasons and details to support your answer.

[Topic3. Argument — Communication]
Do you agree or disagree with the following statement? Television has destroyed communication among friends and family. Use specific reasons and examples to support your opinion.

Repetition Group : Topic 2 – Topic 2 – Topic 3
Non-Repetition Group : Topic 1 – Topic 2 – Topic 3
Effects of Rehearsal on ESL Learners’ Responses to Test Tasks

Betsy Lavolette
Second Language Studies Program
Michigan State University
betsy@msu.edu

Second language (L2) testing is often stressful for test takers, especially when they take high-stakes tests such as the TOEFL. On the iBT TOEFL independent speaking questions, for example, test takers are given 15 seconds to prepare to respond to a prompt (Educational Testing Service, 2008) and have only one chance to record their answer, which puts test takers under great pressure. However, the necessity of this stressful situation has not been empirically validated; that is, the effects of giving learners planning opportunities are unclear.

The major types of planning that are generally distinguished are pretask planning and within-task planning (Ellis, 2005). Pretask planning is subdivided into strategic planning and rehearsal, and within-task planning is subdivided into pressured and unpressured planning (Ellis, 2005). Strategic planning is operationalized by giving learners time (often about 5 to 10 minutes) to plan before being asked to perform a task, while rehearsal is operationalized as repetition of a task. Within-task planning is regarded as pressured when a short time limit is set for the performance of a task.

Definition of a Task
The studies reviewed below (and much of the other research on tasks) present their tasks such as video narration (Gass, Mackey, Alvarez-Torres, & Fernández-García, 1999) and poster presentation (Lynch & Maclean, 2000) without any discussion of whether these meet the definition of a task. While tasks have been defined in many different ways, the one that I will adopt here is that of Ellis (2003, pp. 9–10). The key features of a task are (a) a primary focus on meaning, (b) an information gap that learners must communicate to fill, (c) no specification of what linguistic resources must be used to fulfill the task, and (d) a communicative outcome beyond a gratuitous display of language. The poster presentation described by Lynch and Maclean (2000), which simulates the presentation of a poster at an academic conference, fits this definition well, while one could argue that the video narration of Gass et al. (1999) fails on criteria b and d, given that the researchers have also seen the video. However, another way to view the issue is to consider activities on sliding scale, where some possess more task-like qualities than others. The task used in the present study, in which learners responded to a TOEFL prompt, is similar to the video narration in that it has some task-like qualities (i.e., it meets criteria
a and c). While not all of the tasks in the studies reviewed below meet all of the criteria, I will refer to them as tasks.

Tasks can also be divided into types based on their purposes. Pedagogic tasks are primarily aimed at enabling student learning, while test tasks are intended to assess student learning.

**Pedagogic Tasks**

*Strategic Planning.*

Researchers have extensively examined strategic planning for pedagogic speaking tasks, with the general result that this type of planning improves fluency and complexity, but the results are mixed regarding its effect on accuracy (e.g., Crookes, 1989; Foster & Skehan, 1996, 1999; Mehnert, 1998; Ortega, 1999; Wendel, 1998).

*Rehearsal.*

Researchers have also investigated how rehearsal affects learners’ task performance. Gass, Mackey, Alvarez-Torres, and Fernández-García (1999) asked L2 learners of Spanish to narrate the action in silent video clips in their L2. The participants were divided into three groups: one group (Same Content group) saw the same clip at Times 1 through 3, then a new clip at Time 4; a second group (Different Content group) saw a different clip each of the four times; and a control group saw two different clips, one at Time 1 and one at Time 4. These viewing times were separated by 3 to 4 days and took place in a laboratory setting. Overall, no significant differences were found between the groups on complexity, accuracy, or fluency, with one exception. In terms of lexical sophistication, the Same Content group used more lower frequency words at Time 4 compared to Time 1 than did the Different Content group or the control group. These results indicate that the task repetition had little effect on overall performance, fluency, and accuracy and that only the effect on lexical sophistication carried over to a new task.

Lynch and Maclean (2000) studied two learners at very different levels of English as they repeated a poster presentation task six times during one class session, without instructor feedback. They found that the advanced learner improved in pronunciation accuracy and lexicogrammatical performance. The beginning learner gained in syntactic, lexicogrammatical, and phonological accuracy from the first performance to the last.

Bygate and Samuda (2005) experimentally examined the effects of task repetition, but with a much longer interval of 10 weeks between the first and second performances. They analyzed the performances of 48 learners of English who narrated short cartoon clips and found an effect on fluency and complexity.

The results of these three studies found rather different effects, possibly because of the varied tasks, intervals between repetitions, numbers of repetitions, interactivity (monologic or dialogic), settings (classroom or laboratory), and proficiency levels of the participants. One characteristic that these tasks have in common is that they were all conceived as pedagogical in nature. Test tasks, on the other hand, may affect learner output in a different way, as considered in the next section.
Test Tasks

Despite the many pedagogical studies above that found effects of strategic planning, in testing situations, almost no effects of this type of planning have been found. Iwashita, McNamara, and Elder (2001) found no effect of 3.5 versus 0.5 minutes of pre-task planning time on speaking test task ratings or complexity, accuracy, and fluency (CAF) determined by close analysis of the resulting discourse. Similarly, Wigglesworth and Elder (2010) found no difference on ratings or CAF for learners who were given 15 seconds, 1 minute and 15 seconds, or 2 minutes and 15 seconds of planning time before performing speaking test tasks.

No studies have looked at the effects of rehearsal on learners’ responses to test tasks. Speculatively, rehearsal of a test task may differ from that of a pedagogic task because of the differing levels of stress in each situation and the learner’s focus. The learner taking a test may feel much greater pressure to give his or her best performance than a learner in a normal classroom situation. In addition, to the learner taking a test, “best performance” may mean an accurate performance, rather than one that includes the most complex or fluent language that he or she can produce.

Current Study

Learners are put under tremendous stress to respond quickly to prompts on tests such as the TOEFL, without any evidence that pretask planning and opportunities for rehearsal make a difference in their scores or CAF. No studies have looked at how learners respond to test tasks when they are given the opportunity to rehearse, in addition to pretask planning time. Therefore, in this study, I ask,

What effect does rehearsing the response to a prompt have on ESL learners’ CAF and holistic ratings if they repeat the response once? More than once?

Bygate and Samuda (2005), drawing on Levelt's (1989) model of speech production, claimed that task rehearsal has an effect on both conceptualization (planning the propositions to be expressed) and formulation (choosing the lexical and grammatical elements needed to express the propositions), and thus improves complexity and fluency:

Hence on the second occasion, formulation is likely to be speedier and more accurate. In addition to these influences, clearly the improvement of speed and accuracy of the conceptualization processes outlined above is likely to make more capacity available at the formulation level. If we think of repetition as enabling a second ‘draft’, then task repetition involves targeting improvement not just of the draft (i.e., the language produced) but of the actual drafting process. That is, task repetition can have an impact on the processing, and not just on the product. (Bygate & Samuda, 2005, p. 45)

However, given that Iwashita et al. (2001) and Wigglesworth and Elder (2010) found no effect for strategic planning in testing situations, Bygate and Samuda's (2005) reasoning may not hold in the present
case. I predict that, rather than affecting complexity and fluency, test task repetition will lead to improvement of only accuracy. As Iwashita et al. (2001) suggested, the testing situation itself may alter the focus of the learners:

In a test, where tasks are carried out alone in a computer-mediated environment and hence lack an interactive dimension, the cognitive focus may be on display, and this may alter the relation between task characteristics and language output. For example, a focus on accuracy may be paramount in the testing situation regardless of the conditions under which the task is performed, and this in turn may affect the fluency and complexity of candidates' speech. Delivery may be halting whether the task is easy or difficult, because the candidates are focusing primarily on correctness. The lack of complexity in candidates' production may likewise be due to their anxiety about how their speech is being evaluated, making them reluctant to venture beyond what they know how to say properly even when the task conditions allow for this. (Iwashita et al., 2001, p. 431; emphasis added)

Thus, given that learners are likely to be focused on accuracy in a testing situation, the trade-off hypothesis (e.g., Foster & Skehan, 1996), which was developed to account for differences in CAF under different task conditions, predicts that the learners will have fewer attentional resources to devote to complexity and fluency. When learners repeat the test tasks in the current study, I predict that accuracy will improve, but not complexity and fluency.

Method
Participants
Thirty-nine English-language learners enrolled at the Michigan State University English Language Center (ELC) participated in this study outside of their normal class hours. Nineteen of the learners were in Level 3 classes and 14 were in Level 4 classes in the intensive English program, 3 were in English for academic purposes courses (already enrolled in the university), and the levels of 3 were unknown. The learners were invited to participate in the study by their teachers, at the request of the researcher. They were told that they would practice for the independent speaking portion of the iBT TOEFL test, and they received extra credit in their classes for participating.

Materials
I used TOEFL iBT Test Independent Speaking prompts to elicit speech from the participants. For this analysis, only responses to one prompt will be considered:

Some college students choose to take courses in a variety of subject areas in order to get a broad education. Others choose to focus on a single subject area in order to have a deeper understanding of that area. Which approach to course selection do you think is better for students and why? (ETS, 2006, p. 230)
This prompt was originally published by ETS in a TOEFL preparation book. The rubric that was used to score the recordings was published on the Internet by ETS; it is the same one used in the iBT TOEFL. An opinion and background questionnaire was also administered to the learners; see the Appendix.

Procedure

The learners came to a computer lab outside of their normal class time. The researcher explained the study and demonstrated the technology that was used. Then, the learners recorded three audio and three video speech samples in response to TOEFL iBT Test Independent Speaking prompts using Audio and Video Dropboxes (created by the Center for Language Education and Research at Michigan State University, http://clear.msu.edu/clear/index.php), with the order of the prompts and technology counterbalanced. The first sample of each mode (one audio and one video) was used for the learners to practice using the technology and was not analyzed. The learners read the prompts on the computer screen, then made as many recording attempts as they liked. They were given no time restrictions on pretask planning, note taking, or rehearsal attempts, and they were also given no guidance. All rehearsal attempts were recorded. The time for the response was limited to 45 seconds, as in the TOEFL test, which served to limit their online planning time. After completing the recordings, the learners filled out a questionnaire on their opinions of the two modes of recording and demographic information. All of the learners received extra credit in their courses for participating in the research.

Analysis

Only the responses to one of the prompts in the audio mode were considered. This prompt (see above) was chosen because of the convenient numbers of participants who made one or multiple recordings: Nine of the participants made two or more audio recordings for this prompt, and 9 of the participants made only one.

Ten trained raters rated the recordings holistically, using the TOEFL rubric. Two raters, including the researcher, evaluated CAF. Grammatical complexity was evaluated using the amount of subordination, operationalized as the total number of clauses divided by the total number of AS units. Lexical complexity was measured using the number of words outside the most frequent 1000 English words divided by the total number of words in the response. Accuracy was evaluated using the general measure of error-free clauses and the specific measure of target-like use of finite verb phrases. Fluency was evaluated by dividing the number of syllables in the pruned speech by the total time allotted (45 seconds).

Discussion and Conclusion

Limitations

The learners decided themselves whether they would make multiple recording attempts or just one, which means that they, in effect, self-selected whether they were in the experimental group or control group. This makes the division into the groups nonrandom and limits the conclusions that can be drawn from this
study. In addition, the learners self-selected the amount of time that they spent on pretask planning, making that variable uncontrolled. It is also possible that the learners rehearsed their performances before they actually began recording, although I did not see evidence of this.

Another limitation is that the students were not asked how they oriented to the task. That is, although the task is viewed as a test by the researcher, the learners themselves may not have treated it that way. Given that previous results showed a difference in learners’ performance based on the task being a test or not, the results need to be interpreted with caution.

Testing Implications

If the results are as anticipated, the learners in this study will improve on accuracy but not on holistic ratings, fluency, or complexity when they repeat the task. The learners who repeat the task will also have higher accuracy than learners who do not repeat the task. On the other hand, if results are not as anticipated, an alternative explanation is as follows: A learner focus on accuracy when performing a test task may not lead to gains in accuracy when the task is repeated simply because of the focus on accuracy from the beginning. Instead, it is possible that in subsequent performances, the learners will switch their attention to other aspects of production.

If learners were allowed to rehearse on the real TOEFL speaking test, they might similarly increase their accuracy (or alternatively, other aspects of production), which would change the outcome of the test by biasing for the best (Fox, 2004; Swain, 1983). The test could still be regarded as “fair” in the sense that all test takers would be given the same opportunity to rehearse. I regard this as a positive change because, as Swain (1983) claimed,

[I]f the testee does well, then it can be said with some confidence that the learner can do what is expected of him or her when given the opportunity. However, if the testee does not do well, then it is not clear whether this occurs because the testee cannot do what is expected, or is prevented from doing it because of other distracting factors, or whatever. (p. 141)

The chief advantage to allowing this is to reduce the stressfulness of a very high-stakes exam, which may allow test takers to better show their full capabilities.
References


Appendix

**Learner Background Questionnaires**

1. Age in years:
2. Gender
   - Male
   - Female
   - Other
3. What is your first language?
   - Mandarin (Chinese)
   - Cantonese (Chinese)
   - Korean
   - Arabic
   - Japanese
   - Spanish
   - Other __________
4. How many years have you been studying English?
5. How long have you been living in the US or another English-speaking country?
6. Have you taken the TOEFL before? If so, which version did you take most recently?
   - iBT
   - PBT
   - CBT
7. What was your most recent TOEFL score?
Student attitudes toward accentedness of native and non-native speaking English teachers

Laura Ballard
Second Language Studies Program
Michigan State University
ballar94@msu.edu

The goal for this study was to examine participants’ familiarity with specific accents, whether participants were able to identify if a speaker was a native speaker (NS) or a non-native speaker (NNS), and what accent the speaker had. I also examined how the participants rated speakers on four Likert-scales of comprehensibility, intelligibility, accentedness, and acceptability as a teacher (the four dependent variables). I included 38 NS and 94 NNS participants from a range of first-language backgrounds. The participants listened to three NSs (Midwestern U.S., Southern U.S., and British) and two NNSs (Chinese and Albanian) and completed the identification and Likert-scale tasks outlined above. Results showed that NNSs were significantly less able than NSs to identify a speaker’s nativeness and accent. Results revealed that familiarity with an accent correlated with comprehensibility and acceptability as a teacher. For familiar accents, familiarity was a significant predictor of the participant ratings on the four dependent variables, though the predicted changes in ratings were small. Overall, participants had generally positive attitudes toward NNSETs; in relation to acceptability as a teacher, accent was the least influential of the dependent variables. I conclude by arguing that students should be exposed to a range of different accents, as familiarity with an accent facilitates comprehension. These findings also challenge current language center hiring practices that exclude NNSETs from jobs based on a non-native status; this study supports the notion that administrators should hire English language teachers based on professional credentials, and not based on accent.

In the wake of a recent law passed in Arizona that is based on the No Child Left Behind Act (see Title III, Sec. 3116, c), which threatens the jobs of non-native speaking English teachers (NNSET), the qualifications of both native speaking English teachers (NSET) and NNESTs have come into question. Some argue that this law, which mandates that K-12 English language teachers must be proficient speakers of English (which Arizona has partially interpreted as having a near-native accent, or a least not having a noticeable foreign accent) is clearly employment discrimination (Blum & Johnson, 2012) that has been justified by the myth that English language learners (ELL) will not be able to acquire native-like English from non-native speakers (NNS). In a study conducted by Moussu (2010), around 60% of English
language program administrators surveyed admitted that the primary factor they consider when hiring new teachers is being a native speaker (NS). Because of this continued practice of what could be viewed as discrimination against NNSETs, many in the English-as-a-second-language (ESL) field have sought to defend NNSETs by highlighting their natural strengths as English teachers. Among these qualities are the ability to serve as successful role models for ELLs, a more comprehensive knowledge of English grammar, a better grasp of learning strategies, an anticipation of potential language problems, and a greater empathy for student needs and problems (Hertel & Sunderman, 2009; Moussu, 2010). Furthermore, when there is a match between the teacher’s and students’ L1, the teacher may be able to use the L1 as a beneficial teaching tool: some have argued that some use of the L1 is beneficial (Hertel & Sunderman, 2009; Marian, Blumenfeld, & Boukrina, 2008; Thoms, Liao, & Szustak, 2005; Park, 2012). Despite this realization of what NNSTs bring to the classroom, language center administrators are still more likely to hire NSETs. This practice is based on the idea that NSETs are better and more qualified, which is a belief Philipson (1992) termed the native speaker fallacy. This fallacy has been perpetuated in a way that impacts education policies, as it has in Arizona.

Many in the Teaching English to Speakers of Other Languages (TESOL) profession argue that administrators should hire teachers on the basis of linguistic ability and professional skill (Holliday, 2008; Moussu, 2010; Todd & Pojanapunya, 2008), and not based on accent, as many fear the new law in Arizona is advocating. This employment discrimination disregards the benefits of exposing language learners to a variety of cultures and accents and, whether intended or not, will promote “linguistic sterilization” (Blum & Johnson, 2012, p. 175). This may be especially concerning when one considers the growth in the number of non-native English speakers around the world. Canagarajah (2005) estimated that almost 80% of English speakers worldwide are non-native speakers. Because of this, notions of World English, English as an international language (EIL), and English as a lingua franca (ELF) are becoming more prevalent (Llurda, 2004; Blum & Johnson, 2012; Moussu, 2010). Providing persuasive evidence for this argument is the inclusion of accented, non-native English on some versions of the International English Language Testing System (IELTS) listening section (Harding, 2012; Moussu, 2010) and the recent creation of the Vienna-Oxford ELF Corpus, a corpus that compiles English language samples from the non-native English speaking world (Llurda, 2004).

English language teachers, both teachers of ESL and English as a foreign language (EFL), need to help their students become aware of the varieties of English used across the globe. This is because non-native English, such as EIL and ELF, are becoming stabilized languages and are used frequently in educational, political, and economic contexts.

Alptekin (2002) suggested that NNESTs have a strong role in filling the gaps between the native-speaker and
nonnative-speaker reality:

With the increasing establishment of English as the world lingua franca, non-native speakers will be in optimal positions to lead their students into the realm of EIL. Teachers of EIL should incorporate instructional materials and activities rooted in local as well as international contexts that are familiar and relevant to language learners’ lives. (p. 318)

Blum and Johnson (2012), among others, have argued that there is no longer such a thing as pure or “proper” English. Even among English speakers who learned English as their native language, such as people from Nigeria, Guyana, Australia, and Singapore, there is no universal standard, especially in terms of accent and lexicon. Considering this, it seems rather odd to enforce a law under which teachers can only be employed if they have a local English accent. This law seems quite unreasonable, especially in a collegiate ESL context where the majority of international students plan to return to their own country where many expect to use English for international business purposes. Not only this, but international students attending large American universities will likely also encounter a number of international teaching assistants during their time as undergraduates. As I will review shortly, having an exposure to nonnative speakers throughout their ELL career would benefit learners and aid in their comprehension with other non-native speakers.

Despite these benefits for ELLs, the question still remains about the legitimacy of the Arizona law and the hesitancy to hire NNESTs. To help answer this question, in this study I investigate the issue of non-native-teacher accents from the perspective of the learner by measuring student judgments of comprehensibility and intelligibility of native and non-native English accents in relation to students’ perceptions of a speaker’s acceptability as a teacher. Before describing the study in depth, I highlight past research on accent, student perceptions, and native and non-native teachers.

**Literature Review**

Alford and Strother (1990) conducted a study to capture native and non-native students’ attitudes toward regional U.S. accents. They showed that non-native speakers were able to detect differences between the accents, which is contradictory to research by Derwing and Munro (1997), Mousssu (2010), and Scales et al. (2006), who demonstrated that ELLs had difficulty in distinguishing between different accents, whether they be native accents, non-native accents, or a mix of the two. This was also evidenced when students reported that they wished to acquire an American accent, yet they could not identify one, even after having lived in the U.S. for a number of months or years (Scales, Wennerstrom, Richard, & Wu, 2006). Scales et al. suggested that a desire to have an American accent, paired with an inability to identify one, possibly reflects an idealized conception of what a native accent really is (Timmis, 2002). If students find it difficult
to identify a native accent, perhaps their attitudes toward NNSETs based on pronunciation are invalid.

Despite learners’ inability to distinguish between NS and NNS accents, in Munro, Derwing, and Morton (2006), the participants were able to reliably rate speakers’ intelligibility, comprehensibility, and accentedness. This was true across groups of students with different L1 backgrounds, including a group of native speakers, and their assessment of speakers was surprisingly comparable. In Scale et al.’s (2006) study, students tended to like an accent more if they rated it as more comprehensible. This suggested that students are concerned with listening comprehension, which likely plays a role in student perceptions of teachers.

A number of studies (Gass & Varonis, 1984; Winke, Gass, & Myford, 2013; Harding, 2012; Derwing & Munro, 1997) have also looked at how familiarity relates to comprehension and listener bias. Gass and Varonis looked at four aspects of familiarity: topic familiarity, different L1 background familiarity, same L1 background familiarity, and speaker familiarity. Results showed that being familiar with a topic, a specific speaker, and non-native speech in general facilitates comprehension. They also showed that being familiar with a particular non-native accent facilitates comprehension of other non-native speakers with the same background. This last finding was also consistent with Winke, Gass, and Myford’s (2013) findings, showing that raters who were familiar with a speaker’s L1 were better able to orient themselves to the speaker’s language. In the same vein, for testing purposes, Harding (2012) looked at accent familiarity among ELL students of the same L1 as the speaker. Results showed that students were sometimes able to comprehend more from someone who was speaking English and had the same L1, calling this an *L1 advantage*. These finding suggest that the more familiar a student is with their teacher and his or her linguistic background, and even his or her L1, the easier it will be for them to understand their teacher. Moussu’s (2010) findings (that student ratings of their teachers (both NSETs and NNESTs) over the course of a semester improved) provide additional evidence that exposure to an accent and a particular speaker facilitates comprehension. This is important when thinking about students who have NNESTs who do have an accent. These studies suggest that it is only a matter of time before a student will become comfortable with their teacher’s speech, if they wish to do so.

In the case that a teacher, or speaker, has a strong accent, it does not necessarily mean that the speaker is incomprehensible. This conclusion was consistent with Derwing and Munro’s findings (1997, 2005) in which participants rated accent most harshly, followed by comprehensibility, and then intelligibility. In other words, though the participants judged a speaker to have a heavy accent, the raters admitted to being able to understand the speaker. The researchers suggested that the harsh ratings for accent and comprehensibility might have been due to the fact that a strong accent requires more processing time, and in turn is more difficult to understand.
It’s important to consider why accent might be important in terms of the NSET versus NNSET debate. According to Moussu (2010), a teacher’s L1 is only one of the many variables that affect how students view their teacher. In her study, Moussu collected data at the beginning and end of a semester and found that students’ views of their NNSETs became more positive over the semester. This finding was corroborated by Todd and Pojanapunya (2009) in their findings that, while students made explicit indications that they preferred NSETs, their implicit attitudes suggested that the students value NNSETs equally and that they had “warm feelings” toward NNSETs and not toward NSETs. Hertel and Sunderman (2009) also found that, while students prefer NSET for pronunciation purposes, students recognized NNESTs’ ability to give grammatical rules and explain vocabulary. Among other factors that affect students’ perceptions toward teachers are the students’ expected grades, the students’ majors, and a teacher’s country of origin (Moussu, 2010).

Drawing upon the conclusions from these studies, it is clear that more research needs to be done that will clarify the issues that have seen mixed results. In this study, I seek to add to the literature on accentedness, student attitudes, and NESTs and NNSETs. Especially considering the current policies coming into play regarding English-teachers’ accents (i.e., the law in Arizona that prohibits heavily accented NNSs from teaching), I hope to shed light on students’ attitudes toward such teachers.

Research questions

1a. Are ESL learners able to distinguish between native and non-native speakers of English?

1b. If so, can the ESL learners identify the speakers’ accents?

2a. Do students who are familiar with a particular accent rate speakers differently in terms of comprehensibility, intelligibility, accentedness, and acceptability as a teacher?

2b. If so, is how they rate comprehensibility, intelligibility, and accentedness related to their attitudes about acceptability as an English teacher?

Method

Participants

As shown in Table 1 and Table 2 (available in Appendix A), 132 individuals participated in this study. I recruited these participants from nine classes at a large Midwestern university. The participants included 93 international students, the majority of whom were enrolled in part-time or full-time ESL classes in three proficiency levels: level 2 (n=26), level 4 (n=25), and level 5 (n=30). Of these international students, Chinese was the largest L1 group (n=56), followed by Arabic (n=19), Korean (n=7), and several other languages spoken by three or fewer participants, from this point on referred to as other (n=12). I also recruited a control group (n=38) of American L1 English speakers from two
undergraduate and one graduate-level linguistics courses. All participants ranged in age from 18 to 51 years, with a mean of 22.4. With the exception of the control group, the participants were recruited in intact classes. Teachers of these intact classes volunteered one class period for the purpose of this study, and students were required to participate as an in-class activity, but were able to release or decline the release of their survey information. The control group was recruited from an undergraduate linguistics course and was offered extra credit to participate in the survey.

Table 1

*Participants by L1 Group*

<table>
<thead>
<tr>
<th></th>
<th>Chinese</th>
<th>Arabic</th>
<th>Korean</th>
<th>English</th>
<th>Other</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NNS of English</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>42</td>
<td>11</td>
<td>4</td>
<td>0</td>
<td>5</td>
<td>62</td>
</tr>
<tr>
<td>Female</td>
<td>14</td>
<td>8</td>
<td>3</td>
<td>0</td>
<td>7</td>
<td>32</td>
</tr>
<tr>
<td>Total</td>
<td>56</td>
<td>19</td>
<td>7</td>
<td>0</td>
<td>12</td>
<td>94</td>
</tr>
<tr>
<td><strong>NS of English</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Female</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>33</td>
<td>0</td>
<td>33</td>
</tr>
<tr>
<td>Total</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>38</td>
<td>0</td>
<td>38</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>56</td>
<td>19</td>
<td>7</td>
<td>38</td>
<td>12</td>
<td>132</td>
</tr>
</tbody>
</table>

Table 2

*Participants by Program*

<table>
<thead>
<tr>
<th></th>
<th>B.A.</th>
<th>Master's</th>
<th>Ph.D.</th>
<th>VIPP</th>
<th>ESL2</th>
<th>ESL4</th>
<th>ESL5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NNS of English</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>19</td>
<td>17</td>
<td>23</td>
<td>62</td>
</tr>
<tr>
<td>Female</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>7</td>
<td>8</td>
<td>7</td>
<td>32</td>
</tr>
<tr>
<td>Total</td>
<td>1</td>
<td>7</td>
<td>2</td>
<td>3</td>
<td>26</td>
<td>25</td>
<td>30</td>
<td>94</td>
</tr>
<tr>
<td><strong>NS of English</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Female</td>
<td>31</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>33</td>
</tr>
<tr>
<td>Total</td>
<td>35</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>38</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>36</td>
<td>10</td>
<td>2</td>
<td>3</td>
<td>26</td>
<td>25</td>
<td>30</td>
<td>132</td>
</tr>
</tbody>
</table>

*Note.* VIPP = visiting professionals and scholars; ESL = students who are currently enrolled in English-as-a-second-language classes; ESL2 = intermediate low; ESL4 = intermediate high; ESL5 = advanced low.
Materials
In this study, I had one main instrument, a multi-faceted, web-based survey. The survey had four main components: a background questionnaire, Likert scale questions to be answered in response to listening tasks, a dictation task, and an indication of the speakers’ native status and accent.

Background questionnaire. I adapted the questions on the background questionnaire from Harding (2011). The questions addressed the participants’ exposure to other languages and to English spoken by a variety of native and non-native speakers. The participants used the questionnaire to indicate their preferred accent or accents when listening to English and their general attitude about their own native-like-accent goals.

Audio files and Likert-scale questions. The audio files, which native speakers (American Midwestern from Michigan, British from Northern English, American Southern from Alabama, and Nigerian from Lagos) and non-native speakers (Chinese, Albanian, Malagasy, and Italian) recorded, are of three kinds: a lecture about a familiar topic that ranged between 20 and 24 seconds (classroom expectations given at an assembly that each ESL student attends); a lecture on an unfamiliar topic that ranged between 20 and 24 seconds (pottery making, adapted from Sueyoshi & Haridson (2005); and unrelated sentences for transcription that are similar in syntax, topic, and length that ranged between 5 and 7 seconds.

Dictation Task. The final group of audio-recorded sentences served as an audio prompt for a series of sentence-dictation tasks that could serve as an alternate measure of intelligibility, as has been done previously in research studies on accent (Derwing & Munro, 1997; Munro et al., 2006). Because of the finding that topic familiarity influences comprehension (Derwing & Munro, 1997; Gass & Varonis, 1984), I have included familiar and unfamiliar topics.

Nativeness and Accent. The final component of the survey required participants to indicate whether they thought the speaker was a NS or a NNS. They were also required to guess the speakers’ accents. For a sample of a listening page for one speaker in the survey, see Appendix A.

Procedure
The first group of participants in this study was an intact graduate-level linguistics class. During class, the participants met in a computer lab to complete the survey. Each student had their own computer and a set of headphones for the listening portion of the survey. The participants first completed a familiarization task that is similar to the listening portion. In this familiarization task (and the listening section), participants responded to 9-point Likert-scale questions after hearing the lectures. The scales measured, for each participant, his or her estimation of the speaker’s intelligibility, comprehensibility, accentedness, and acceptability as a teacher, with the participant assuming that the lecturers were ESL teachers. These Likert scales were presented in a random order to eliminate bias or an ordering effect. Following Munro et al.’s (2006) definitions, I define intelligibility as “the extent to which a
speaker’s utterance is actually understood” (p. 112), comprehensibility as “the listener’s estimation of difficulty in understanding an utterance” (p. 112), and accentedness as “the degree to which the pronunciation of an utterance sounds different from an expected production pattern” (p. 112).

Next, the participants completed the background questionnaire and the main listening task with Likert-scale questions (as described above). The participants listened to each lecture from the familiar topic and unfamiliar topic groups twice. Each time participants listened to a lecture, they answered two of the Likert-scale questions. Thus, each participant listened to each lecture twice and, in total, answered four Likert-scale questions per lecture. For counterbalancing purposes, the participants listened to the speakers and answered the Likert-scale questions at random (which SurveyMonkey automatically arranges). Each participant listened to one familiar and one unfamiliar lecture from each speaker, totaling sixteen lectures.

Next, the participants completed a dictation task, which served as an alternate measure of intelligibility. In this task, each participant completed one dictation per lecture. This task was presented at the end of each Likert-scale evaluation for each speaker. The participant listened to the recorded sentence once and then transcribed the sentence, to the best of their ability, using a text box.

Finally, the participants indicated their best guess about the nativeness and accent of the given speaker. The first round of participants, however, completed an expanded survey that included eight speakers, four native and four non-native. After this group of participants completed the survey, they gave feedback to the researcher, and from a combination of the initial data and the feedback collected, this allowed me to remove any speakers that either had no perceivable accent or that had such a strong accent that they were unintelligible. At this point, I narrowed the listening portion to include only the American Midwestern (from Michigan), American Southern (from Alabama), British (from Cumbria, Northern England), Chinese and Albanian speakers.

Next, each of the remaining classes, with the exception of one undergraduate linguistics course, completed the survey in a computer lab during class time. They followed the same procedure; however, they did not give feedback, and they completed the shortened survey of only five speakers (10 total lectures). For one undergraduate linguistics class, I oriented the students to the survey and gave instructions during class time, but the students completed the survey on their own time.

Analysis

To gain an overall picture of the participant population, I calculated descriptive statistics that included nationality, age, sex, program enrollment, and L1. Then to answer the first research question, I calculated descriptive statistics to determine whether participants accurately identified the speakers’ accents and accurately identified whether the speaker was a NS or a NNS. To see if NS participants and NNS participants performed differently in their identification of speakers,
I calculated an independent samples $t$-test. To answer the second research question, I calculated Spearman’s correlations for ordinal data using students’ self-reported familiarity with an L1 and their actual rating of comprehensibility, intelligibility, accentedness, and acceptability as a teacher for corresponding L1 speakers. I also calculated simple linear regressions to determine if familiarity with an L1 predicted ratings for comprehensibility, intelligibility, accentedness, and acceptability as a teacher.

**Results**

To answer the first research question, whether ESL learners are able to distinguish between NSs and NNSs, I calculated descriptive statistics of participants’ ratings of speakers as NS or NNS. These statistics revealed that NS participants, in general, were more able than NNS participants to distinguish between NSs and NNSs. On average, NS participants correctly identified the speakers’ native status 91% of the time, as compared to NNS participants, who correctly identified speakers only 68% of the time. Next, to see if these differences were significant for each speaker, I calculated an independent samples $t$ test. Results from the $t$ test revealed that in all cases, except for the British speaker, NS participants were significantly better at determining the native status of the speaker. This was true for both the familiar and unfamiliar lectures given by each speaker. Table 3 reports these results. To answer the latter part of the first research question, whether ESL learners are able to identify a speaker’s accent, I calculated descriptive statistics on participants’ guesses of speakers’ accents. Table 4 is a summary of these results. Raw scores for correct and incorrect answers revealed that, for every lecture, NSs were better able to identify the accent of the speaker, and in six of the ten lectures, NSs were at least 17% more accurate in their identification of the speakers’ accent. I then calculated independent samples $t$-tests to see if the differences observed between participant groups were significant. Results revealed that differences between NS and NNS guesses were significant for the familiar Midwestern lecture, $t(130) = -12.52$, $p = .000$, the unfamiliar Midwestern lecture, $t(130) = -8.05$, $p = .000$, the familiar southern lecture, $t(130) = -11.60$, $p = .000$, the unfamiliar southern lecture, $t(130) = -3.99$, $p = .000$, the familiar British lecture $t(130) = -5.07$, $p = .000$, and the unfamiliar Albanian lecture, $t(130) = -2.57$, $p = .014$. Results for the averages were also significantly different between NSs and NNSs on the familiar lectures ($t(130) = 5.1139$, $p = .0001$) and on the unfamiliar lectures ($t(130) = 3.0147$, $p = .0031$). For full results, see Table 5 and Table 6.
Table 3
Correct Identification of Speakers as NS or NNS by Participant Group (NS and NNS)

| Participant Group | Familiar | | | | | | Unfamiliar | | | | |
|-------------------|----------|---|---|---|---|---|---|---|---|---|---|---|
|                   | NS       | NNS | T  | Df | Sig. | NS | NNS | t  | df | Sig. | | |
| Midwestern        | 100%     | 61% | -7.649 | 91 | 0.00* | 100% | 69% | -6.44 | 93 | 0.00* | | |
| Southern          | 100%     | 83% | -4.368 | 93 | 0.00* | 97% | 77% | -4.06 | 130 | 0.00* | | |
| British           | 68%      | 58% | -1.124 | 72 | 0.27 | 60% | 60% | -0.03 | 68 | 0.97 | | |
| Chinese           | 97%      | 86% | -2.538 | 127 | 0.01* | 100% | 86% | -3.86 | 93 | 0.00* | | |
| Albanian          | 95%      | 57% | -5.960 | 127 | 0.00* | 92% | 42% | -7.37 | 117 | 0.00* | | |

Note. An asterisk denotes a significant p value of less than .05. The percentages reflect the mean for each participant group, which are also the percentage of participants who correctly identified the speaker as a NS or NNS.

Table 4
Summary of Correct Identification of Speaker Accent by Participant Groups

| Participant Group | NNS of English | | | | | | NS of English | | | | |
|-------------------|----------------|---|---|---|---|---|---|---|---|---|---|---|
|                   |                | NS | NNS | T  | Df | Sig. | NS | NNS | t  | df | Sig. | | |
| Midwestern        |                | 38% | 100% | | | | | | | | | |
| Familiar          |                | 33% | 89% | | | | | | | | | |
| Unfamiliar        |                | 27% | 63% | | | | | | | | | |
| Southern          |                | 12% | 87% | | | | | | | | | |
| Familiar          |                | 27% | 63% | | | | | | | | | |
| Unfamiliar        |                | 38% | 45% | | | | | | | | | |
| British           |                | 30% | 74% | | | | | | | | | |
| Familiar          |                | 38% | 45% | | | | | | | | | |
| Unfamiliar        |                | 34% | 39% | | | | | | | | | |
| Chinese           |                | 56% | 45% | | | | | | | | | |
| Familiar          |                | 34% | 39% | | | | | | | | | |
| Unfamiliar        |                | 2% | 11% | | | | | | | | | |
| Albanian          |                | 3% | 21% | | | | | | | | | |
| Familiar          |                | 2% | 11% | | | | | | | | | |
| Unfamiliar        |                | 3% | 21% | | | | | | | | | |
| Average           |                | 28% | 63% | | | | | | | | | |
| Familiar          |                | 27% | 52% | | | | | | | | | |
| Unfamiliar        |                | 27% | 52% | | | | | | | | | |

Note. Bolding across pairs denotes a significant difference between NS and NNS ratings. Familiar and unfamiliar refer to topic.
Table 5

<table>
<thead>
<tr>
<th>Lecture</th>
<th>Participant Group</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Midwestern</td>
<td>NNS</td>
<td>94</td>
<td>.372</td>
<td>.4860</td>
<td>.0501</td>
</tr>
<tr>
<td></td>
<td>NS</td>
<td>38</td>
<td>1.000</td>
<td>.0000</td>
<td>.0000</td>
</tr>
<tr>
<td>Midwest</td>
<td>NNS</td>
<td>94</td>
<td>.330</td>
<td>.4727</td>
<td>.0488</td>
</tr>
<tr>
<td></td>
<td>NS</td>
<td>38</td>
<td>.895</td>
<td>.3110</td>
<td>.0505</td>
</tr>
<tr>
<td>Southern</td>
<td>NNS</td>
<td>94</td>
<td>.117</td>
<td>.3232</td>
<td>.0333</td>
</tr>
<tr>
<td></td>
<td>NS</td>
<td>38</td>
<td>.868</td>
<td>.3426</td>
<td>.0556</td>
</tr>
<tr>
<td>Southern</td>
<td>NNS</td>
<td>94</td>
<td>.266</td>
<td>.4442</td>
<td>.0458</td>
</tr>
<tr>
<td></td>
<td>NS</td>
<td>38</td>
<td>.632</td>
<td>.4889</td>
<td>.0793</td>
</tr>
<tr>
<td>British</td>
<td>NNS</td>
<td>94</td>
<td>.298</td>
<td>.4598</td>
<td>.0474</td>
</tr>
<tr>
<td></td>
<td>NS</td>
<td>38</td>
<td>.737</td>
<td>.4463</td>
<td>.0724</td>
</tr>
<tr>
<td>British</td>
<td>NNS</td>
<td>94</td>
<td>.372</td>
<td>.4860</td>
<td>.0501</td>
</tr>
<tr>
<td></td>
<td>NS</td>
<td>38</td>
<td>.447</td>
<td>.5039</td>
<td>.0817</td>
</tr>
<tr>
<td>Chinese</td>
<td>NNS</td>
<td>94</td>
<td>.553</td>
<td>.4998</td>
<td>.0516</td>
</tr>
<tr>
<td></td>
<td>NS</td>
<td>38</td>
<td>.447</td>
<td>.5039</td>
<td>.0817</td>
</tr>
<tr>
<td>Chinese</td>
<td>NNS</td>
<td>94</td>
<td>.340</td>
<td>.4764</td>
<td>.0491</td>
</tr>
<tr>
<td></td>
<td>NS</td>
<td>38</td>
<td>.395</td>
<td>.4954</td>
<td>.0804</td>
</tr>
<tr>
<td>Albanian</td>
<td>NNS</td>
<td>94</td>
<td>.021</td>
<td>.1451</td>
<td>.0150</td>
</tr>
<tr>
<td></td>
<td>NS</td>
<td>38</td>
<td>.105</td>
<td>.3110</td>
<td>.0505</td>
</tr>
<tr>
<td>Albanian</td>
<td>NNS</td>
<td>94</td>
<td>.032</td>
<td>.1767</td>
<td>.0182</td>
</tr>
<tr>
<td></td>
<td>NS</td>
<td>38</td>
<td>.221</td>
<td>.4132</td>
<td>.0670</td>
</tr>
</tbody>
</table>

*Note.* The mean can be read as the percentage of participants that correctly guessed the accent of the speaker, where 1.00 equals 100% of the participants guessing correctly, and .00 equals 0% of the participants guessing correctly. Bolding denotes that NS and NNS participants’ guesses were significantly different, as noted in Table 4.
Table 6

Independent Samples T- Test between NS and NNS Participants for Correct L1 Accent Guesses

<table>
<thead>
<tr>
<th></th>
<th>t</th>
<th>df</th>
<th>Sig. (2-tailed)</th>
<th>Mean Difference</th>
<th>Std. Error Difference</th>
<th>95% Confidence Interval of the Difference</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mid-western</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Familiar</td>
<td>-12.52a</td>
<td>93.00</td>
<td>0.00*</td>
<td>-0.63</td>
<td>0.05013</td>
<td>-0.73, -0.53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unfamiliar</td>
<td>-8.05 a</td>
<td>102.72</td>
<td>0.00*</td>
<td>-0.57</td>
<td>0.07016</td>
<td>-0.70, -0.43</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Southern</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Familiar</td>
<td>-11.89</td>
<td>13</td>
<td>0.00*</td>
<td>-0.75</td>
<td>0.06321</td>
<td>-0.88, -0.63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unfamiliar</td>
<td>-3.99 a</td>
<td>63.03</td>
<td>0.00*</td>
<td>-0.37</td>
<td>0.09159</td>
<td>-0.55, -0.18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>British</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Familiar</td>
<td>-5.01</td>
<td>130</td>
<td>0.00*</td>
<td>-0.44</td>
<td>0.08765</td>
<td>-0.61, -0.27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unfamiliar</td>
<td>-0.80</td>
<td>130</td>
<td>0.43</td>
<td>-0.08</td>
<td>0.09442</td>
<td>-0.26, 0.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chinese</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Familiar</td>
<td>1.10</td>
<td>130</td>
<td>0.27</td>
<td>0.11</td>
<td>0.09631</td>
<td>-0.08, 0.30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unfamiliar</td>
<td>-0.59</td>
<td>130</td>
<td>0.56</td>
<td>-0.05</td>
<td>0.09263</td>
<td>-0.24, 0.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Albanian</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Familiar</td>
<td>-1.60 a</td>
<td>43.66</td>
<td>0.12</td>
<td>-0.08</td>
<td>0.05263</td>
<td>-0.19, 0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unfamiliar</td>
<td>-2.57 a</td>
<td>42.58</td>
<td>0.01*</td>
<td>-0.18</td>
<td>0.06946</td>
<td>-0.32, -0.04</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. An asterisk denotes statistically significant differences between NS and NNS participant guesses. a = equal variance was not found, thus the t and p values reflect the Levene’s test adjustment for that violation.

Looking further into this question, I ran descriptive statistics for percentage of correct NNS ratings on nativeness and L1 accent by program level to uncover any differences in rating based on a participant’s English proficiency level. Here, I was interested in looking specifically at students currently enrolled in English language courses (ESL2- intermediate low, ESL4- intermediate high, and ESL5- advanced low) and master’s students (TOEFL score of 100+). I did not include NNS participants from the VIPP or B.A. level because their proficiency level was not available. Results show that there were differences between these groups. These findings are graphically represented in Figure 1. These numbers show a clear trend; when proficiency level increases, ability to identify nativeness and accent increase.

The only exception to this trend that I found was when participants listened to the unfamiliar lecture and rated the speakers’ nativeness. In this case, the scores for participants in ESL2, ESL4, and ESL5 were different than expected; ESL2 was higher than expected and ESL5 was lower than expected.
Figure 1. Percentage of correct identification of nativeness and accent by NNS participants by program level.

Figure 2. Percentage of correct identification of nativeness and accent by NNS and NS participants.
To answer the second research question, whether students who are familiar with a particular accent rate speakers differently in terms of the four variables, I began by calculating Spearman’s correlations to uncover significant relationships between students’ reported familiarity with an L1 and their actual rating of the speakers’ comprehensibility, intelligibility, accentedness, and acceptability as a teacher, as shown in Table 7. The results show that the correlations between participants’ self-reported familiarity with an accent and their actual ratings of the four variables were, at best, moderate (my benchmark for moderate being a coefficient ranging between .3 and .5, as outlined in Field (2009)). The strongest correlation was between Midwestern familiarity and comprehensibility of the familiar Midwestern lecture, rs = .44, p < .001, followed by British familiarity and intelligibility of the unfamiliar British lecture, rs = .41, p < .001, and Midwestern familiarity and acceptability as a teacher of the familiar Midwestern lecture, rs = .40, p < .001. All other coefficients were less than .39. It is worthwhile, however, to note that, across the board, every correlation with Albanian was less than .10 and was not statistically significant. With all other accents, most (if not all) of the correlations for each individual accents ranged between .10 and .39. In addition, Midwestern familiarity was significantly correlated with every variable for both the familiar and unfamiliar lectures. Importantly, in terms of categorical variable correlation, familiarity correlated with comprehensibility and acceptability for both lectures (with the exception of Albanian), whereas accentedness and intelligibility revealed mixed results.
Table 7

Spearman’s Correlations between Participants’ Reported Accent Familiarity and Participants’ Actual Ratings of Four Variables for Corresponding Speakers

<table>
<thead>
<tr>
<th>Familiar Lecture</th>
<th>Unfamiliar Lecture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Midwestern</td>
<td></td>
</tr>
<tr>
<td>Familiarity</td>
<td>-.27**</td>
</tr>
<tr>
<td>British</td>
<td></td>
</tr>
<tr>
<td>Familiarity</td>
<td>-.29**</td>
</tr>
<tr>
<td>Southern</td>
<td></td>
</tr>
<tr>
<td>Familiarity</td>
<td>-.18*</td>
</tr>
<tr>
<td>Chinese</td>
<td></td>
</tr>
<tr>
<td>Familiarity</td>
<td>-.22*</td>
</tr>
<tr>
<td>Albanian</td>
<td></td>
</tr>
<tr>
<td>Familiarity</td>
<td>-.02</td>
</tr>
</tbody>
</table>

Note. *Correlation is significant at the 0.05 level (2-tailed). ** Correlation is significant at the 0.01 level (2-tailed). Bolded coefficients are greater than or equal to .3.

To obtain a deeper understanding of the answer to this research question, I also calculated simple linear regressions to see if familiarity with a particular accent predicted participants’ ratings on the four variables when listening to a speaker of the accent. Within each regression, I wanted to see if familiarity with an accent (the independent variables: Midwestern, etc.) was associated with the four dependent variables (comprehensibility, accentedness, intelligibility, and acceptability as a teacher) and if so, to what extent. Results showed that, for both the familiar and unfamiliar Chinese lectures, familiarity with Chinese was a significant predictor of ratings given on all four variables. With the exception of accentedness, this pattern was also true of the British lectures. With the exception of comprehensibility in the familiar lecture, familiarity with the Midwestern accent was a significant predictor of all of the variables in both of the Midwestern lectures as well. For Albanian and Southern, however, the results were quite different. For Albanian, familiarity with the accent was not a significant predictor for any of the variables in either lecture. Similarly, the Southern lectures did not show significant predictions, with the exception of comprehensibility and acceptability for the unfamiliar lecture.

The strongest predictions between familiarity with an accent and ratings of the four variables can been seen for the unfamiliar British lecture (comprehensibility, B = -.378; accentedness, B = .177; intelligibility, B = -.395; and acceptability as a teacher, B = -.475), the unfamiliar Chinese lecture (comprehensibility, B = -.418.; accentedness, B = -.322; intelligibility, B = -.308; and acceptability as a teacher, B = -.288), the familiar British lecture (comprehensibility, B = -.448; accentedness, B = -.302; intelligibility, B = -.298; and acceptability as a teacher, B = -.198), and the familiar Chinese lecture (comprehensibility, B = -.443; accentedness, B = -.303; intelligibility, B = -.293; and acceptability as a teacher, B = -.193).
.383; and acceptability as a teacher, $B = -.406$), and the unfamiliar Midwestern lecture (comprehensibility, $B = -.449$; accentedness, $B = .375$; intelligibility, $B = -.379$; and acceptability as a teacher, $B = -.513$). Here the $B$ value, or the regression gradient, indicates the predicted movement on the rating scale of the dependent variables as they relate to the rating given for familiarity. In other words, if a participant moves one point up the familiarity scale (e.g. from 2 to 3), then the $B$ value predicts that, in the case of acceptability as an English teacher in the unfamiliar Midwestern rating, the rating will move negatively (toward a lower number, or toward “acceptable”) on the scale by 51% of one point (i.e. the rating for acceptability would move half of a point toward “acceptable”).

To answer the second half of the second research question, I calculated Spearman’s correlations to see if there were significant relationships between participants’ ratings of acceptability as a teacher with the other three variables. Results revealed that, across the board for both familiar and unfamiliar lectures, acceptability as a teacher showed, at the very least, a small correlation for those that were noted as significant (which includes 68 of the 70 correlations), with the correlation coefficients ranging from $r_s = .19$ to $r_s = .86$. For accentedness of both lectures, 12 of the 20 ratings were moderately (.3 to .5) to highly (.5 and above) correlated, with correlation coefficients ranging from $r_s = .32$ to $r_s = .67$. Notably, neither lecture for the British speaker had a moderate or stronger correlation (i.e., these correlations were weak, at best). Also, for the Albanian and Southern lectures, either one or both of the lectures showed small correlations. For intelligibility, 18 of the 20 ratings were moderately to highly correlated, with the correlation coefficients ranging from $r_s = .32$ to $r_s = .84$. The only lectures that did not show this trend were the unfamiliar Chinese and Albanian lectures. Finally, comparing acceptability ratings between the familiar and unfamiliar lectures, the correlation coefficients were moderate to strong across the board, ranging from $r_s = .49$ to $r_s = .63$.

**Discussion**

In this study, I sought to gain a greater understanding of student attitudes toward NSETs and NNSETs by testing their accent perception ability and by assessing their reactions to a variety of native and non-native accents. The first research question was as follows: 1a) Are ESL learners able to distinguish between native and non-native speakers of English? 1b) If so, could the ESL learners identify the speakers’ accents? Answering these questions gives insight into students’ ability to differentiate between English accents (including native and non-native accents), and also helps to shape the interpretation of the results of the second research question, which focuses on participants’ attitudes about each type of English accent as related to their exposure to a given accent. The second research question is as follows: 2a) Do students who are familiar with a particular accent rate speakers differently in terms of comprehensibility, intelligibility, accentedness, and acceptability as a teacher? 2b) If so, is how they rate related to their
attitudes about acceptability as an English teacher?

As shown in Figure 2, NNS participants were usually able (68% of the time) to correctly identify a speaker as being native or non-native. Notably, however, NNS participants were significantly less able to make these identifications than NS participants, who made correct identifications 91% of the time. In looking at the correct identification of accent, however, these percentages dropped; NNS were only able to correctly identify an accent 28% of the time, and NS, 57%. This was different from the findings of Derwing and Munro (1997), who found that NNSs were able to correctly identify an accent 52% of the time. In their study, participants had a forced choice between the four accents, which likely made it easier to identify the accents and could explain the higher percentage of correct identification. Similar to the current study, in Derwing and Munro’s study, a sizable portion of incorrect guesses were given to languages or accents from the same language family (e.g. Chinese and Japanese, Spanish and Italian). This could account for a portion of the incorrect guesses, but what other factors could explain this low rate of correct identification on the part of both NS and NNS participants?

One possible explanation is discussed in Moussu’s (2010) study of student attitudes. She conducted a beginning and end-of-the-semester survey of ESL student attitudes toward their NSETs and NNSETs. A review of the latter survey revealed that, even after spending an entire semester with a teacher (and having lived in the U.S. for a semester or more), some students could not correctly identify whether their own teacher was a NSET or a NNSET. The students were clearly familiar with their teacher’s accent (and the American accent spoken in their greater community), so factors other than familiarity with the accent were influencing their identification of the teachers’ native status. Moussu hypothesized that this misidentification could have been related to the teachers’ appearances or presentation styles. In the current study, as in Moussu’s study, many of the students had spent a semester or more in the U.S. and would undoubtedly have spent a great deal of time being exposed to the local accent (in this case, the Midwestern accent). However, Moussu’s explanation would not fit the current situation because the current participants lacked familiarity with the speaker and lacked visual cues that might influence their decisions.

Another possible explanation for this difference was given in the Scales et al. (2006) study, in which 62% of the ELL participants indicated that they wanted to acquire a native-like accent, yet only 29% of the students were able to identify an American accent. Similarly, in the current study, 40% of the ELLs indicated wanting to acquire a native-like accent, but only 30% of learners were able to identify the native English speakers (Midwestern, Southern, and British). As Scales et al. wrote, there was a “mismatch between these learners’ own accent goals and their ability to perceive accents” (p. 728). This mismatch between an idealized accent by a student and the lack of ability to identify the real accent could partially account for the significant differences seen between NSs and NNSs.
participants in the current study.

A number of the factors cited in previous research, such as language families, visual cues, teaching style, and a distorted perception of a native accent, could potentially account for the inability of some participants to correctly identify a speakers’ nativeness (whether they are a native or non-native speaker) and their specific accent. I hypothesize, however, that there is an additional factor that could account for this. As shown in Figure 1, there is a clear trend that shows that as proficiency increases, the ability to distinguish between native and non-native accents and the ability to identify specific accents also increases. As a student becomes more proficient in English, their capacity to notice nuances in pronunciation is also likely to increase, which would explain why more proficient ESL speakers are more adept at detecting accent differences.

In the second research question, I asked whether students who are familiar with a particular L1 rate speakers differently in terms of intelligibility, comprehensibility, and accentedness, and acceptability as a teacher. Perhaps the most strongly evidenced claim that I can make based on the correlation results, provided in Table 7, is that familiarity with an accent is significantly related to a student’s judgment of how easy it is to understand a speaker of that accent and a student’s judgment about the speaker’s acceptability as a teacher. Both of these variables (comprehensibility and acceptability as a teacher) were positively correlated with familiarity; in other words, the more familiar a participant was with an accent, the easier it was to understand the speaker, and the more acceptable the speaker was as a teacher. This relationship between familiarity and comprehensibility aligns with results by Gass and Varonis (1984), who clearly showed that familiarity facilitates comprehension. This observation is further supported by Winke, Gass and Myford (2013), who demonstrated that accent familiarity contributes to comprehension.

Looking deeper into the relationship between comprehensibility and acceptability as a teacher is also telling. In looking at the correlations between these two factors, every correlation was significant, and was moderate to strong, with a coefficient ranging between .341 and .858. The relationship between intelligibility and acceptability as a teacher patterned similarly, with coefficients ranging between .282 and .839. These positive correlations are not surprising, considering that the easier it is to understand a teacher (or, in terms of intelligibility, the more of a teacher’s speech a student is able to understand), the more likely a student is to rate the teacher as acceptable as a teacher. Similar results were also born out in Scales et al.’s (2006) study, where they found that comprehension was a high priority for students, and the students tended to prefer an accent (or teacher) if the teacher’s accent was easy to understand (p. 725).

The results for the linear regressions were also telling. While familiarity with an accent was not a significant predictor of ratings on the four variables in every case, it was overwhelmingly a significant predictor of the four variables for the Midwestern, British, and Chinese accents. Where these
were significant predictions, however, the effect sizes were small. In other words, a one-point movement on the familiarity scale would, at most, predict a half-a-point movement on the scales for the four variables. For example, if a student rated themselves as being very familiar with an accent (5), then they would be likely to move 2.5 points down the scale for acceptability as a teacher (toward acceptable, 1). Considering that the four variables are ranked on a 9-point scale, this type of .5 movement seems rather inconsequential. Putting this in classroom terms, if two students of the same proficiency level are assigned to a classroom with a NNSET whose L1 is Chinese, and one student is very familiar with a Chinese accent and the other student is not, their ability to understand the teacher may only be mildly different, as is the case with their attitude toward the acceptability of the teacher, based on accent alone.

This classroom scenario is especially interesting in light of Moussu’s (2010) findings: in her study, students’ attitudes toward their NNSETs became more positive over the semester. This scenario is also supported by Gass and Varonis (1984), who found that familiarity with a speaker, with non-native speech in general, and with a non-native accent in particular, all facilitate comprehension. As seen in the current study, comprehension and acceptability as a teacher are positively correlated, meaning that (drawing on Moussu’s findings) the more a student interacts with their NNSET teacher, (and also drawing on Gass and Varonis,) the easier it will be to understand their teacher, and (drawing on the current study,) the more positively the students will view their teacher. While this may be true on paper, we know that ESL classes are not a vacuum devoid of other factors that would influence a student’s attitude toward their teacher; yet the evidence is nonetheless compelling in suggesting that the current study showed that students’ judgments of the strength of a teacher’s accent influences the students’ view on whether they find the teacher acceptable as a teacher. In terms of results, I found that accentedness and acceptability as a teacher correlated (small to moderate) with the exception of the British accent. However, as discussed in the previous paragraph, even if a teacher was viewed as having a moderate to strong accent, their students may gradually understand more and more of the teacher’s speech, and the students will understand the teacher with greater ease over time.

Alford and Strother (1990), who asked NS and NNS participants to listen to speech samples and rate them on bipolar adjectives scales (e.g. professional/non-professional, intelligent/unintelligent, etc.), found that ELLs were able to detect differences in accents. These were extrapolated from differences in their expressed attitudes toward the speakers of different accents. In other words, students used different adjectives to describe different accents, and the researchers interpreted this as evidence of the students being able to differentiate between accents. These results also clearly showed that they did have different attitudes toward accented English. Though the ELL participants in the current study were largely unable to overtly
decipher a speaker’s nativeness and accent, like Alford and Strother’s findings, the students’ varied ratings of acceptability as a teacher for different speakers does evidence differences in attitudes based on accent.

In general, there were some inconsistencies in the data that mostly revolved around the Albanian accent and the Southern accent. One explanation for the variations found in relation to these accents is that, for Albanian, the vast majority of the participants had no exposure to Albanian and could not place Albania on a map. This was true for NS and NNS participants alike. Because this language is relatively novel to the participants, I assume that some participants were blindly guessing accents, or that some were making educated guesses, drawing from other semantically related or phonologically related languages (i.e. Indo-European languages, particularly Balto-Slavic or Germanic languages). This could explain the large variance in results for the Albanian accent.

For the southern accent, during data collection, many of the NNS students asked what a southern accent was (what it sounded like), or if it was the same as a South American accent. This confusion or misunderstanding could likely have been experienced by many more participants, and thus could be a possible explanation for the variation in the Southern accent data. Whereas ratings for Albanian and Southern varied widely in the data, the results for Midwestern, British, and Chinese were undoubtedly more consistent. This consistency could be attributed to a greater general knowledge or general true familiarity with these cultures and accents, whereas ratings for Albanian and Southern could have been compromised by the misunderstanding and general unawareness that was noted during data collection. Another important note about familiarity is that participants rated themselves as being most familiar with the British accent. This could be explained by the largest NNS groups represented in their participant population, Chinese and Saudi. The educational systems in these countries largely train their students in Received Pronunciation, also known as standard British English, which accounts for the high reporting of British familiarity. I would contend, however, that many of these students only have a superficial exposure to Received Pronunciation. For example, students in China often are strong readers, but lack oral communicative competence due to the lack of focus on speaking and listening in the Chinese classroom. For both of these groups (Chinese and Saudi), the majority have never studied abroad, or even traveled outside of their countries, and have very limited experience with spoken English. I would, therefore, assume that students have overreported their familiarity with the British accent, and have perhaps underreported their familiarity with the Midwestern accent (especially since the majority of the NNSs had lived in the Midwest for 6 months or more at the time of data collection).

Inconsistencies aside, participants’ ratings have clearly shown that accentedness is not synonymous with being an unacceptable teacher. Ratings for accentedness were higher than ratings for acceptability as a teacher, shown in Figure 3.
Student attitudes toward accentedness

On average, participant ratings for accentedness of the familiar and unfamiliar lectures were 3.74 and 3.70, respectively (where 1 is “no accent” and 9 is “very strong accent”), and ratings for acceptability as a teacher were 2.23 and 2.60, respectively (where 1 is “acceptable” and 9 is “not acceptable”). The ratings for acceptability were an average of 1.5 and 1.0 points lower than participants’ ratings of accentedness. These ratings, relative to one another, reveal a general acceptance and positive attitude toward teachers with any type of accent.

Students’ acceptance of accented teachers provides evidence that challenges the assumptions of language center administrators, such as in Moussu’s (2010) study, which reported that one of the major factors in hiring English language teachers is their native speaker status. Many administrators claim that they continue in these hiring practices because of “native speaker demand” on the part of students. Results in this study, however, provide insight into student attitudes toward NNSETs, based on their accents, which shows that student attitudes toward NNSETs are generally positive.

Assuming that students do have a positive attitude toward NNSETs, Clarke and Garrett (2004) provided further evidence in support of Moussu’s findings; Clarke and Garrett found that listeners quickly adapt to new accents, and applying this to the current study, students will adapt to the accents of their teachers. In their study, they investigated how long it takes NSs to adapt to foreign-accented English speech, and found that NSs can adapt to accented speech in as little as one minute. Presumably, the speech processing load placed on NNSs as they listen to English would be much heavier, and would therefore take NNSs longer to adapt to accented speech. This, however, provides added insight into Moussu’s finding, as discussed earlier, and could be a possible explanation of why students’ attitudes toward their NNSETs became more positive as students were exposed to their teachers’ speech over the semester. This process is also supported by the current research, which shows that familiarity, comprehensibility, and students’ attitudes about a teacher’ acceptability are all related.

Conclusion

In this study, I have sought to extend previous research on accentedness, native versus non-native speaking English teachers, and ESL student attitudes. As previously outlined, ESL students were usually able to distinguish between native and non-native speakers. However, they were generally not able to identify a speaker’s accent. Some factors that might explain this inability to identify speakers are a confusion of languages in a language family, misconceptions concerning native accents, and a student’s proficiency level. Results further uncovered that familiarity is correlated with comprehensibility and acceptability as a teacher. In this study, comprehensibility, intelligibility, and acceptability were highly correlated with one another. For truly familiar accents, familiarity was a significant predictor of participant ratings on the four variables (comprehensibility, intelligibility, accentedness, and acceptability as a teacher),
Student attitudes toward accentedness

though the predicted changes in ratings were small. The current research shows that students had generally positive attitudes toward NNSETs; based on previous research, students’ attitudes toward their NNSETs only become more positive over time, and students may quickly perceptually adapt to their accented teachers, if they are willing to do so (Derwing, Rossiter, & Munro, 2002). Finally, this research has shown that accentedness does not, in the minds of students, translate to unacceptability as a teacher.

Like any study, this research has its limitations. One is that the survey was long. The fatigue that students experienced from answering the same questions again and again could have resulted in a lack of concentration. A second limitation is that the audio clips embedded in the survey were not limited to a single play. This means that students could have listened to any audio file as many times as they wanted (though they were instructed to listen only once), which could have given some students an unequal advantage. Finally, during one round of data collection, the participants experienced technical difficulties with the survey, which frustrated the participants. This also could have influenced how they approached the survey and the answers that they gave.

I believe that the implications of these findings are most pointed for the hiring practices (or in the case of Arizona, firing practices) of schools and language centers; as the results have shown, student attitudes toward NNSETs were generally positive, and even when students indicated that a teacher had a pronounced accent, the students still demonstrated a positive attitude toward the teachers by rating them as acceptable. This shows that, from a student’s perspective, other factors may more heavily influence a student’s attitude toward a teacher and whether or not a teacher is acceptable. As TESOL² also argued, teachers should be hired on the basis of their professional credentials, such as professionalism, teaching experience, and English proficiency as a whole.

Educators should also consider what is best for students. Given the nature of diversity at any university, the rapid growth of EIL and ELF, the overpowering number of non-native English speakers, and the purposes for which many students are learning English, it is in a university or a program’s best interest to expose students to a wide range of global Englishes. As shown in this study, as in many previous studies, a familiarity with an accent leads to greater comprehension or intelligibility of language spoken with that accent (Gass & Varonis, 1984; Winke, Gass, & Myford, 2013; Harding, 2012; Derwing & Munro, 1997) and this would benefit students as they encounter various native and non-native speakers of English throughout the span of their academic, personal, and professional lives.

Directions for Future Studies

Even though the results of this study show quantitatively that students’ attitudes towards their teachers are not largely based on accent, qualitative research needs to be done to corroborate this finding. In addition, though accent adaptability research has been done with NSs, further work that
investigates the amount of time it takes ELLs to adapt to a new accent could provide insight into accent dynamics. This type of information could corroborate evidence based on comprehensibility, intelligibility, and attitudes toward encountering new accents.

\footnote{The TOEFL, or the Test of English as a Foreign Language, is an academic English proficiency test administered by Educational Testing Services. For more information, see the TOEFL website at www.ets.org/toefl}

\footnote{Teachers of English to Speakers of Other Languages, Inc. For more information, see www.tesol.org.}
References


Appendix A: Sample Listening Page

Sample Listening Tasks from Survey

Listen to this audio file ONE TIME.

1. How easy was it to understand this speaker? (please choose one number):

   1 (Very easy to understand)  2  3  4  5  6  7  8  9 (Very difficult to understand)

   Comprehensibility  
   1  2  3  4  5  6  7  8

2. How strong is this person's accent? (please choose one number):

   1 (No accent)  2  3  4  5  6  7  8  9 (Very strong accent)

   Accentedness  
   1  2  3  4  5  6  7  8

Listen to this audio file ONE TIME.

3. How much of this speech did you understand? (please choose one):

   100% (I understood everything)  88%  76%  54%  42%  30%  19%  10%  0% (I didn't understand anything)

   Intelligibility  
   1  2  3  4  5  6  7  8
4. How acceptable is this speaker as an English teacher? (please choose one number):

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acceptability as an English teacher</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5. Listen to this audio file ONE TIME. Then, in the box below, type exactly what you hear.

[Audio playback]

6. I think that this person is a:

- native English speaker
- non-native English speaker

7. I think that this speaker's nationality is:

- Albanian
- American (Midwestern)
- American (Southern)
- Australian
- British
- Chinese
- French
- I don't know
- Indian
- Japanese
- Malagasy
- Nigerian
- Spanish